

nature



FORM FINDS FUNCTION

Enzyme activity predicts
from protein structure

ORGANICS ON MARS

Will Phoenix challenge Viking data?

THE SAN ANDREAS FAULT

Tale raises the friction

CANCER AND AGING

Two sides of the same coin

NATURELIFE
Life & Health



Mbeki's mistake

South Africa's government has removed the minister most closely associated with public discussion of the country's HIV epidemic. But it must stand by its promises to implement a fresh AIDS strategy.

The dismissal on 8 August of South Africa's deputy health minister, Nozizwe Madlala-Routledge, sends out an extremely negative message about how seriously the country is taking its monumental AIDS crisis (see page 739).

Madlala-Routledge was a driving force behind South Africa's first realistic national AIDS strategy, which sets out proposals to cut infection rates, improve diagnosis and treat the estimated 5.5 million South Africans already infected with HIV. The plan was endorsed by the South African National AIDS Council on 30 April (see *Nature* 447, 1; 2007).

President Thabo Mbeki claims that he asked for Madlala-Routledge's resignation because she travelled to an AIDS vaccine conference in Spain earlier in the year without receiving the required permission to make the trip. But no one believes this petty transgression to be the real issue. The fact of the matter is that Madlala-Routledge's direct and honest approach to AIDS and other health challenges had placed her on a direct collision course with both Mbeki and his health minister, Manto Tshabalala-Msimang.

Sidelined earlier this year by medical problems, Tshabalala-Msimang has now returned to an active role in government, and is again championing the tragically misguided idea that food products such as beets are more useful for treating AIDS than antiretroviral drugs.

The deputy health minister — who is a substantial political figure in South Africa in her own right — says she thinks a factor in her dismissal was her speaking out on the shocking conditions she found in the maternity ward at Frere Hospital in East London when she visited it last month. Whatever the precise circumstances, it seems clear that Madlala-Routledge is a victim of her own outspokenness — and of the return of her boss, the health minister, to her desk.

Yet that outspokenness is exactly what is required of public-health officials in South Africa right now. The firing is a particularly bitter blow, because the fresh national AIDS strategy had given patients' advocates, scientists and doctors real hope that the nation would at last move from its failed approach of playing down the threat posed

by AIDS. Now, it is by no means clear that the strategy will be implemented in full.

Although he has refrained from speaking out on the topic lately, Mbeki has come close to embracing AIDS 'denialism' — the rejection of the hypothesis that HIV causes AIDS. In South Africa, this is often aligned with claims that antiretroviral drugs are more dangerous than HIV itself. The overall result of this view from the top is that South Africa, despite its relative prosperity, has been slower than other African nations in distributing medicines that would extend the lives of people who have HIV.

Denialism has also infected the wider South African public: in patient surveys, half of the South Africans who first tested positive for HIV in 2005 said that they had not believed themselves to be at risk of contracting HIV, according to UNAIDS. The UN agency also reports that almost a million South Africans who need antiretroviral drugs are not getting them — and that the epidemic in the country is yet to peak.

The dismissal of Madlala-Routledge augers very badly for South Africa's HIV/AIDS response. Under its previous incoherent strategy, the nation's public health has deteriorated. HIV/AIDS spurs epidemics of other diseases, such as extremely drug-resistant tuberculosis. It also undermines the heart of the health system by killing so many health workers.

The new AIDS strategy had signalled that South Africa's leaders were ready to take a new course — to work with patients, scientists, advocates and international organizations to confront HIV's destruction of their country's human and economic resources. The ministry of health has stated that despite the dismissal it will pull out "all stops" to implement the strategy. It is imperative that it does so. Madlala-Routledge's removal was a serious error of judgement; if the strategy now unravels, it will be a calamity for South African public health. ■

"Outspokenness is exactly what is required of public-health officials in South Africa right now."

Division of labour

The European Research Council shouldn't be coy about saying who will get its first set of grants.

The first Europe-wide research agency to distribute funding purely on the basis of scientific merit is working with commendable efficiency. Its officials have just ploughed through more than 9,000 first-stage applications for the inaugural programme of grants and asked 559 of them to submit a complete application. Around half of these shortlisted candidates will eventually win

five-year grants worth up to €400,000 (US\$550,000) per year.

The European Research Council (ERC) has done well to get so far within eight months of its official creation. But it is already facing criticism for its reluctance to reveal the exact distribution of nationalities on the shortlist. The ERC's decision to keep this information to itself for the time being can be read two ways: as a failure to be transparent or as a pragmatic response to a tricky political environment.

The ERC's mission is perhaps unprecedented in the brief history of the European Union (EU). It has to distribute large amounts of European money — building up to €1 billion a year within a few years — to the best research proposals, regardless of nationality or other political criteria. Both the EU member states and the

European parliament have fully signed up to this mission.

Nonetheless, the young agency's leadership can expect to take some political heat if, as is likely, most of its grants go to those EU countries that are already most established scientifically. A comparable dilemma has been encountered in the past by the US National Science Foundation (NSF), an agency that, perhaps more than any other, the ERC seeks to emulate. NSF grants have always flowed disproportionately to certain states, such as Massachusetts and California, where US scientific excellence is most heavily concentrated. The agency has dealt with the political challenge that this presents by publishing reams of relevant data upfront, while developing programmes (at the prompting of Congress) that assist researchers in the states that do less well with their applications. It has done this without compromising its criteria for grant selection.

One of the council's top priorities is to make sure that it establishes a reputation for excellence in its processes. It must do this to win the solid support of European scientists ahead of its first formal evaluation by the EU authorities, which will take place in just two years' time. For now, the council is still negotiating the details of the final EU executive agency within which it will eventually operate. Evaluation of grant proposals, meanwhile, is being overseen by a modest number of staff, most of whom have been seconded from national research agencies.

It is in this fragile context that the ERC is eager to avoid rocking political boats by publishing a national breakdown of who is being considered for its first grants. Instead, it has broken down the shortlist into the groups of nations that joined the EU at different stages of its evolution.

So it has revealed that 45% of the applicants, and 53% of the winners, come from Belgium, Germany, France, Italy, Luxembourg and the Netherlands — the six original members of the European Economic Community, as it was then known. The nine countries that joined after 1973, but before the entry of the former communist states, account for 36% of applications and 27% of the winners. The 12 members who have joined since 2004 did not do so well, putting in 9% of the applications and winning 5%. (Nine 'associated countries', such as Russia and Israel, as well as participants from farther afield account for the rest of the applications.)

Policy-makers might benefit from fuller information about the geographical distribution of both those who apply and those who make the shortlist, if only as a snapshot of how excellence in European science is currently distributed.

And according to its mission statement, the ERC is "committed to providing public information about its activities in a transparent and timely manner". Ultimately, that commitment to transparency will have to override the council's concerns about giving offence.

EU politics, in its complexity and fickleness, is likely to pose challenges for the new research agency at some stage. But Europe needs the ERC to be openly committed to uncompromising selection of the best. Sooner, rather than later, the ERC needs to commit to full publication of data on its selection processes, to defend these processes to the full, and to let the political chips fall where they may. ■

"Ultimately, a commitment to transparency will have to override the European Research Council's concerns about giving offence."

Men [*sic*]

Our 1869 mission statement is out of date.

It was 1833 when the English polymath William Whewell first coined the word 'scientist'. Over subsequent decades, the word gradually replaced such commonly used terms as 'natural philosophers' and 'men of science'.

By the middle of the nineteenth century, this last phrase was already out of date: pioneering women such as Mary Fairfax Somerville and Caroline Herschel were proving their worth as astronomers, mathematicians, botanists and palaeontologists.

The original mission statement of this journal, first printed in *Nature's* second issue on 11 November 1869, was therefore running behind the times when it referred to "Scientific men" — even though, to be fair, the word 'scientist' did not enter general circulation until the end of the nineteenth century. In other respects it is well worded — which is why we print it every week in the Table of Contents.

The statement expresses two purposes for this publication. The first is "to place before the general public the grand results of Scientific Work and Scientific Discovery; and to urge the claims of Science to a more general recognition in Education and in Daily Life". Today this is as important as it has ever been — although members

of the public have important considerations to lay before scientists, and *Nature* reflects them also.

The second thrust was expressed as follows: "to aid Scientific men themselves, by giving early information of all advances made in any branch of Natural knowledge throughout the world, and by affording them an opportunity of discussing the various Scientific questions which arise from time to time."

In printing the statement verbatim every week as we have done, making it clear when it originated, we have hitherto assumed that readers will excuse the wording in the interests of historical integrity. But feedback from readers of both sexes indicates that the phrase, even when cited as a product of its time, causes displeasure. Such signals have been occasional but persistent, and a response is required.

There is a convention within the English language by which writers quoting text can indicate their view that a particular phrase is inappropriate. That is to insert *sic*, a Latin word meaning 'thus', after the phrase — in effect expressing the sentiment 'alas, dear reader, this is what was said'.

This is what we will do in the mission statement from now on. The small, belated change takes place against the vast backdrop of a scientific world where the upper echelons of academia, academies and prestigious awards are still numerically greatly dominated by men, and where outright discrimination can still rear its ugly head (see page 749). In this context, the insertion of a Latin word in a couple of paragraphs may be a tiny step: but it is at least one in the right direction. ■

RESEARCH HIGHLIGHTS

GENETICS

Pace of change

Science 317, 813–815 (2007)

The rate at which the bacterium *Escherichia coli* accumulates beneficial mutations may be 1,000 times higher than previously thought, according to research by a team in Portugal.

Isabel Gordo at the Gulbenkian Science Institute in Oeiras and her colleagues tracked mutation rates in small and large populations of *E. coli*. In large populations, the team found, many mutations that would make the bacteria more fit are in effect 'swept away' by a few, more strongly adaptive mutations. The effect, known as clonal interference, is less pronounced in smaller populations. Failing to account for this effect properly may have led to underestimates of the mutation rate in previous studies. The findings may help to explain how pathogenicity and antibiotic resistance evolve quickly in bacteria.

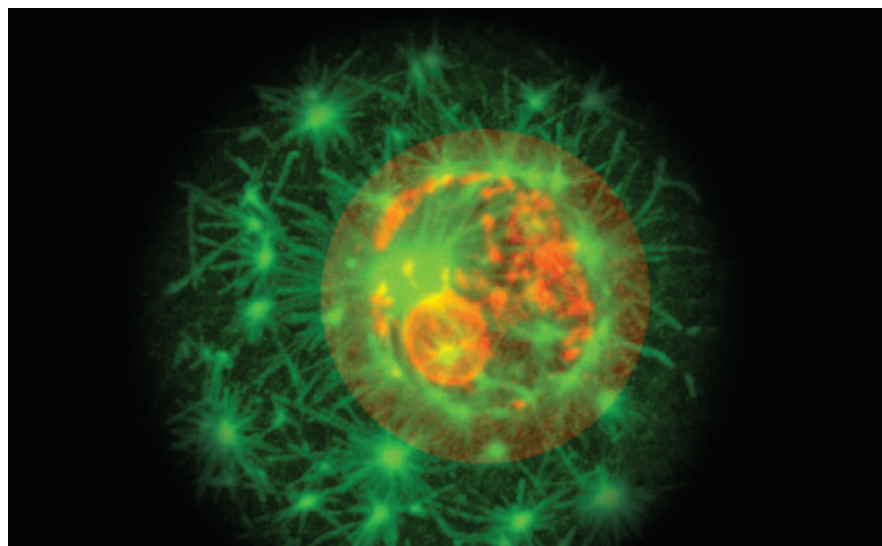
ANIMAL BEHAVIOUR

High turnover

Funct. Ecol. doi:10.1111/j.1365-2435.2007.01321.x (2007)

The nectar-eating bat *Glossophaga soricina* (pictured below) has the fastest metabolism of any mammal studied so far, say Christian Voigt of the Leibniz Institute for Zoo and Wildlife Research in Berlin, Germany, and John Speakman of the University of Aberdeen, UK.

The researchers gave bats of this species food with different levels of carbon isotopes, then measured how quickly the isotopes appeared in the bats' breath. They estimate that the bats get more than three-quarters of their energy directly from food yet still turn



Split in half

Cell 130, 484–498 (2007)

Despite lacking a crucial piece of the scaffolding that helps most cell types to package their chromosomes during cell division, mammalian egg cells manage to halve their number of chromosomes before fertilization. How do they do it?

Jan Ellenberg and Melina Schuh of the European Molecular Biology Laboratory in Heidelberg, Germany, have solved this mystery with the help of high-resolution

images of maturing mouse oocytes.

In non-reproductive cells, a structure called a microtubule spindle guides chromosome segregation. This assembles around two smaller structures known as centrosomes, which do not exist in mouse oocytes. Ellenberg and Schuh found that the spindle in oocytes assembles from more than 80 temporary structures, dubbed microtubule organizing centres (green in picture; chromosomes red), created just before the egg divides.

over half of their body fat in 24 hours.

Humans fuel about three-quarters of their metabolism from stores of fat and glycogen, but it would be impractical for bats to carry large fat stores in flight.

PHYSIOLOGY

Not just a bunch of bones

Cell 130, 456–469 (2007)

The skeleton is more than just scaffolding for the body. New work has revealed that, in mice at least, it regulates aspects of metabolism.

Key to the skeleton's role is a protein called osteocalcin, produced by bone cells known as osteoblasts. Researchers led by Gerard Karsenty at Columbia University in New York found that osteocalcin influences blood sugar levels and fat deposition. Mice lacking this protein showed traits of diabetes, such as insulin resistance. Conversely, mice engineered to be in some senses anti-diabetic, suffering effects including low blood sugar and increased insulin sensitivity, showed reduced symptoms if one allele of the osteocalcin gene was deleted.

The results suggest that osteocalcin

functions as a hormone, making the skeleton an endocrine organ, and hint that it might have a use in the treatment of diabetes.

CHEMISTRY

Frenetic kinetics

Angew. Chem. Int. Edn doi:10.1002/anie.200701326 (2007)

Researchers have developed an apparatus that can rapidly determine reaction parameters, such as the rate constant and activation energy, for catalytic reactions. This should aid the design of more effective catalysts for industrial applications.

Oliver Trapp at the Max Planck Institute for Coal Research in Mülheim, Germany, and his colleagues built a microreactor in which polymer-bound catalysts are deposited on the walls of capillaries that serve as reaction vessels. The polymer doubles as a chromatographic separation material, so that catalysis and product separation are efficiently combined.

In a study of palladium-catalysed hydrogenation reactions, 5,880 reactions were performed in 40 hours. If such an

analysis were carried out by one person using traditional methods it could take weeks.

EARTH SCIENCES

Leaks and flows

Geophys. Res. Lett. **34**, L13612 (2007)

Ocean circulation in the southern hemisphere is characterized by swirling basin-wide currents, known as gyres, contained within the Indian, Pacific and Atlantic Oceans. Now scientists have discovered a 'supergyre' that connects the three basins.

Ken Ridgway and Jeff Dunn of CSIRO Marine and Atmospheric Research in Hobart, Australia, determined flow patterns from temperature and salinity profiles collected since 1950. They found that, in a region south of Tasmania, Pacific waters 'leak' into the Indian Ocean. There is a similar but weaker connection south of Africa between the Indian and Atlantic Oceans.

Models have suggested that the ocean basins are linked, but until now the connecting element had escaped observation. The supergyre seems to be an essential component of the global ocean circulation system.

CELL BIOLOGY

Age: nothing but a pathway

Science **317**, 803–806 and 807–810 (2007)

A molecular pathway that regulates stem cells leads a double life as an usher of old age, two teams have found.

Toren Finkel at the National Heart Lung and Blood Institute in Bethesda, Maryland, and his co-workers examined a strain of mice, known as Klotho, that ages prematurely. The team found that the klotho protein these mice lack keeps the Wnt pathway in check. Wnts are a family of proteins known to stop stem cells from differentiating.

Thomas Rando, at Stanford University School of Medicine in California, and his team noticed the same counterintuitive role for Wnt when they connected the veins and arteries of young and aged mice, so that blood flowed freely between the two. Under these conditions, muscle stem cells in young mice showed signs of fibrosis — a hallmark of old age. Activating Wnt signalling in young mice had the same effect.

ASTRONOMY

Blasts from the past

Astrophys. J. **664**, 1000–1010 (2007)

Short gamma-ray bursts have been a feature of the Universe for longer than some observations have suggested, say Edo Berger of the Observatories of the Carnegie

Institution of Washington in Pasadena, California, and his colleagues.

In seconds, these bursts release energy comparable to the Sun's output over its entire lifetime. Berger's team detected nine such events and estimate that up to two-thirds of this type of burst happened more than six billion years ago. They also estimate higher energies for these ancient bursts than have been reported for more recent bursts.

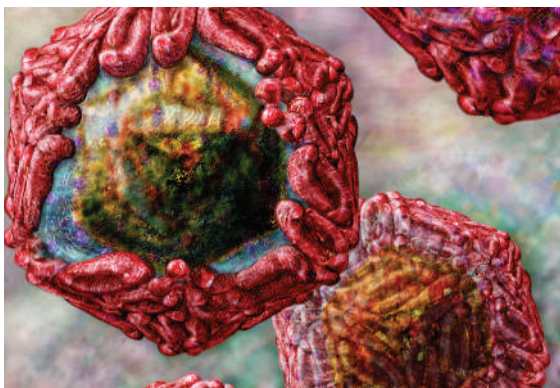
Such observations should help astronomers to work out what causes the explosions: today's best guess is that they result from the merger of two neutron stars, or a neutron star and a black hole.

VIROLOGY

Killer mutation

Nature Genet. doi: 10.1038/ng2097 (2007)

A single mutation may be responsible for the recent increased virulence of the West Nile virus (pictured below). Since 1996, the virus has caused regular outbreaks of severe and even lethal encephalitis in humans. Previously the mosquito-borne, bird-infecting virus had caused only mild infections in humans.



Aaron Brault of the University of California, Davis, and his colleagues analysed the genomes of 21 strains of West Nile virus from around the world. They identified a single base alteration in a gene encoding a helicase enzyme that had arisen independently in at least three strains associated with human outbreaks. Engineering this mutation into a less virulent strain made the virus more lethal to birds, but whether the mutation is associated with increased virulence in humans remains to be determined.

R. KIGHTLEY/SPL

Correction

The Research Highlight 'When dinos hit puberty' (*Nature* **448**, 391; 2007) wrongly implied that modern crocodilians are descended from dinosaurs. Both groups derived from archosaurs.

JOURNAL CLUB

Martyn Poliakoff,
University of Nottingham, UK

A champion of environmentally friendly chemistry encourages attempts to identify reactions ripe to be turned 'green'.

The aim of 'green chemistry' is to make the design, production and use of chemicals more sustainable. This means that, unusually for an academic discipline, industrial implementation is an inherent goal.

Research groups in this field, including mine, strive to reduce waste by identifying selective catalysts, alternative solvents or renewable feedstocks that could lead to new industrial processes.

But how do we choose which reactions to try to green? Some targets are obvious; the reactions are notoriously inefficient. However, many chemical manufacturers are understandably reticent about the shortcomings of their processes.

It was therefore particularly refreshing to find a paper that results from the collaboration of seven pharmaceutical companies and highlights key research areas for green chemists (D. J. C. Constable *et al.* *Green Chem.* **9**, 411–420; 2007). The paper describes several classes of reaction that, if 'greened', would significantly lessen the pharmaceutical industry's effect on the environment.

For example, the paper asks that researchers develop methods to carry out oxidations safely in non-chlorinated solvents (chlorinated solvents are non-flammable but toxic); or to find ways to tame the fearsome reactivity of fluorine so that fluorination occurs selectively.

Another clear message is that new strategies for using solvents could lead to substantial reductions in waste. Could reaction vessels be cleaned out at the end of a process without using organic solvents?

This paper is a great start, but I think the authors have been too conservative. They could have asked for more, such as catalysts that can trigger two or more reactions in sequence. We need really tough challenges to intrigue academic chemists and bring new blood to the task of greening chemistry.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS



Responding to an anthrax scare in Liverpool, UK in 2001.

Not so secure after all

In the movie *28 Days Later* a deadly virus escapes from a British research lab and wreaks havoc across the country. That was fiction, but concerns about lab safety are not.

It is now nearly certain that the foot-and-mouth virus discovered on 3 August in cattle near Guildford, UK, originated at the nearby animal-research facility in Pirbright. The incident seems to have been due to an accidental leak of the virus from either the government-run Institute for Animal Health (IAH) or commercial vaccine manufacturer Merial Animal Health, which share the Pirbright facility. Merial said last week that it “has complete confidence in its safety and environmental protection”. The IAH also says it does not know of any security breaches and is cooperating with the inspectors.

This latest incident highlights the problems that can occur with the security of so-called ‘dual-use’ research — work that could be of use to terrorists as well as to legitimate researchers (see ‘Laboratory lapses’).

Investigations into the foot-and-mouth outbreak are ongoing, but engineering or personnel failure must have been to blame if the

virus escaped from a secure lab, in the opinion of Keith Plumb, a bioprocess engineer at the Institution of Chemical Engineers in London. It could have emerged only through the ventilation system, in waste, or on people, he says. Waste should be sterilized before disposal in the sewers, either by steam or chemicals. Damage to filters in the negative-pressure air system, for example, could have given the virus a possible exit route, says Plumb.

Lab workers are fully covered by a gown, with

only their eyes exposed, and must enter the lab via air-locks. After leaving the lab and removing the gown, researchers must shower to get rid of any contamination that might have occurred. Not taking enough time to shower is another possible exit route for the virus, Plumb says.

“These kinds of breaches happen frequently in labs,” he says, although usually with no serious consequences.

The 2001 anthrax attacks in the United States were a rude awakening for biosecurity, and

I. HODGSON/REUTERS

Laboratory lapses

- **1998** Imperial College London fined for failure to follow health-and-safety rules in a study that created a chimera of hepatitis C and dengue-fever viruses.
- **2001** Anthrax spores sent maliciously through the post in the United States. A laboratory source for the

bacterium was suspected.

- **2003** Thomas Butler of Texas Tech University charged with infringing regulations on the handling of the plague bacterium, including bringing samples into the United States from Tanzania on a plane without declaring them to customs.

- **2003/2004** SARS cases due to laboratory accidents in China, Taiwan and Singapore.

- **2007** Texas A&M University work on ‘select agents’ shut down after failure to report a 2006 incident in which members of staff had been infected with *Brucella* and *Coxiella*. **D.C.**

experts are now warning that biologists need to engage more seriously with the dual-use issue. The prospect of a deliberate release of dangerous biological material is of increasing concern, whether through lab break-ins or by research personnel. The failed terrorist attacks in July in central London and at Glasgow airport seemed even more shocking when it emerged that many of the suspects were doctors. "People can't just walk in off the street into labs, but if there was somebody who wished to do with bioagents what was done with these other things we'd have a job to safeguard against it," says Plumb.

Regulate or be regulated

National and international regulators are currently looking at whether lab safeguards are adequate and how rigorously they are applied. "Biosecurity is probably still a little bit too lax," said Jens Kuhn, a virologist at Harvard Medical School who advises on arms control.

In the United States, the most stringent regulations apply to labs working with 'select agents', which include Ebola virus, saxitoxin (a lethal toxin found in algae), foot-and-mouth-disease virus and bird flu, are required to adhere to a stringent set of rules. Suggestions for site security in the rules include guard services, gated entry, biometric readers and locked storage units. In addition, inventory control and IT systems must be assessed.

However, the US Department of Health and Human Services (DHHS), which is responsible for regulating many dangerous materials, has identified "serious weaknesses" in the security of select agents in university laboratories. These problems were initially identified after a review of 11 universities in 2002. A follow-up investigation of 15 universities in 2004, the results of which were released last year, identified problems with 11 that "could have compromised the ability to safeguard select agents from accidental or intentional loss".

Access controls, such as procedures for issuing electronic keys to sensitive areas, were problematic at six universities. Training for individuals with access to dangerous agents had either not been provided or not been documented at three universities. Inventory and access records were also a problem, with some records incomplete or "difficult to decipher".

Contacted by *Nature*, the DHHS insisted that all the weaknesses had been corrected. But the number of problems identified raises questions about the integrity of labs in general and biology labs in particular.

"It's fair to say that unlike areas such as physics, biologists haven't had that kind of discussion in recent years," says Brian Rappert, a biosecurity

expert at the University of Exeter, UK.

Rappert has been running seminars on biosecurity with Malcolm Dando, professor of international security at the University of Bradford, UK. They found that biosecurity awareness among the 1,600 biologists who attended their seminars was extremely low. "Our regular finding was that very few of them had thought very much about these dual-use issues," says Dando.

This lack of engagement with security may come back to haunt biologists as the trend towards regulation accelerates. Reynolds Salerno, manager of the International Biological Threat Reduction programme at the US government's Sandia National Laboratories in Albuquerque, New Mexico, points out that the international community is beginning to consider the possibility that malicious individuals might acquire dangerous pathogens or toxins from a legitimate laboratory. "The biotechnology community does need to be more involved in helping government establish the risk of dual-use technology and

least provide cogent arguments not to over-regulate these new technologies," he told *Nature*.

Next year, countries including Britain, China, Germany and the United States will revisit the international Biological and Toxin Weapons Convention,

to which more than 150 countries are signatories. "One of the questions is the whole issue of biosafety and oversight, awareness, codes of conduct," says Dando. "We'll see increasing attention from scientific bodies in the run up to that."

Europe is also looking at biosecurity. In July, the European Commission released a consultation paper on bio-preparedness, noting that "in some domains imperfect implementation of safety measures and the existence of security gaps may continue to pose a risk". Proposals on the cards include limiting the number of security-cleared researchers and accredited labs, mandatory courses on biosafety at university level and obliging researchers to adopt a professional code of conduct.

With regulation already a burden, many researchers fear that more rules could be stifling. "The more security you have, the more impaired the research gets. It's definitely very difficult and gets more difficult by the year," says Kuhn. "In the US, in general the people who make these regulations don't consult with the scientists. Whereas physicists have been dealing with these issues for a very long time and have shown research can continue very successfully, biologists just see another hurdle in their grant writing, more paperwork and their papers disappearing. They get panicky."

Daniel Cressey

ZOO NEWS

Croc on the run

Emergency workers in Ukraine are hunting a crocodile named Godzilla that escaped from a circus. Pity the local guinea-pigs: newspaper reports say the furry critters will be used as bait.



SCORECARD



'Liposuction' for pits
Forget antiperspirant, excessive underarm sweat can be reduced to a damp memory by use of a technique that sucks out the sweat glands.



Liposuction is the pits
It's now 'normal' to be obese according to a study that shows American women have become fatter since it's become more socially acceptable to carry extra weight.

A. VAN ZANDBERGEN/LOVELY PLANET IMAGES/GETTY

ON THE RECORD

"Even if it has Velcro on it, you set it aside and within 30 seconds it's gone and you have no idea where it went to. So we'll be playing some treasure hunts."

Teacher-turned-astronaut Barbara Morgan describes the challenges of her trip to the International Space Station, along with her 10 million cinnamon basil seeds.

ROBOT NEWS

So you think you can dance

Researchers have recruited a prancing dancebot to keep traditional dance alive. So far it has learned the *aizu bandaisan* — a Japanese folk dance — but Sidelines wants to see it do the *Time Warp*. It's just a jump to the left...

Sources: BBC, J. Am. Acad. Dermatol., Florida State Univ., NASA, New Scientist, Reuters



SIDELINES

SPECIAL REPORT

Dengue fever climbs the social ladder

Dengue fever was once a disease restricted to poor people in tropical areas. Its resurgence now threatens middle-class urbanites in cities such as Singapore. **Ewen Callaway** asks whether Asia's ever-growing wealth will propel a treatment or vaccine to market.

He was a 25-year-old rock-climber, educated to university level and from a middle-class family in Singapore. But last month, Ian Ng hit the headlines for the worst reason. He became the city-state's fourth and youngest victim of dengue fever this year. Ng was taken to hospital vomiting and feverish on 21 July, but he died from the mosquito-borne virus less than 24 hours later.

The booming cities of southeast Asia are experiencing their worst dengue outbreak in a decade — early rains and unchecked urbanization have promoted the spread of the virus-carrying mosquitoes. Singapore, which almost managed to wipe out the pandemic disease in the 1980s, has already had 5,000 infections this year. In Indonesia, more than 100,000 people have contracted dengue and 1,100 have died, coming close to the figures for all of 2006, according to the World Health Organization (WHO). Malaysia, Cambodia, the Philippines, Myanmar and Thailand are all experiencing a similar surge in the disease. Monsoon flooding across India, Nepal and Bangladesh threatens to worsen the situation by causing an explosion in mosquito populations.

Dengue infects at least 50 million people a year in more than 100 countries, mainly across Asia, Africa and South America. But unlike many diseases ravaging the developing world, dengue cuts across class.

"This is not a disease of the poverty-stricken, rural farmer," agrees Scott Halstead, scientific

director of the Seoul-based Pediatric Dengue Vaccine Initiative (PDVI), a non-profit organization launched in 2003 to drive a vaccine to market. In some developing countries, "the richer you are, the more likely you are to get dengue", he adds. This could be a reason for hope.

Dengue still plagues the poor people in urban areas, who face the brunt of a disease with neither cure nor vaccine. But as Asian economies boom, dengue is starting to interest pharmaceutical companies. "It's been enormously helpful to the field to see the potential market value of a dengue vaccine increase, and that has everything to do with rising incomes in Latin America and Asia," says Bruce Innis, who heads vaccine development at GlaxoSmithKline (GSK). "The reason why GSK and Sanofi-Aventis are in this business is shrewd business judgement," says Halstead.

Andrew Farlow, an economist at the University of Oxford, UK, who is working to determine the potential market for a dengue vaccine, says that there would be a substantial private market for a dengue vaccine in Asia and the Pacific. And the high economic toll that dengue takes in southeast Asia — US\$2.36 billion in the next 10 years according to the World Bank — means that middle- and high-income countries will be willing to pay more.

Spread by the bite of the female *Aedes aegypti* mosquito, dengue is caused by four *Flavivirus* strains. Those infected develop high fever and debilitating joint pain that can leave them bedrid-



den for weeks. Infection with one strain provides lifetime immunity against that strain, but not against the other three. In fact, immunity from one strain boosts the chances of having a much more severe illness, called dengue haemorrhagic fever (DHF), because antibodies against another strain latch onto the virus and speed its entry into host cells. The result is immune overload, as blood vessels leak and the body goes into shock.

Target the hosts

Dengue fever isn't deadly, but DHF kills one in five if patients are not given adequate hospital support. More than 500,000 people — mostly children and infants — contract this more severe form of dengue, which kills 20,000 every year, making it a leading cause of childhood death in many countries. And the numbers are increasing.

Without an effective drug or vaccine, the only way to stem dengue is to go after its host. The mosquitoes flourish in small pools of clean water outside homes — tyres and plastic containers are some of their favourite spots.

Research frontline

In 2003, the Swiss drug firm Novartis launched its Institute for Tropical Diseases (NITD), a public-private partnership based in Singapore with a US\$7 million annual budget to produce drugs to combat dengue.

Several chemicals targeting key viral machinery have shown promise in the lab and NITD hopes to have an antiviral drug ready for

testing by 2009.

In the meantime, NITD is developing tools to quickly distinguish patients with dengue haemorrhagic fever (DHF) from those with dengue fever. For instance, patients with few white blood cells and platelets seem more likely to suffer from DHF.

GlaxoSmithKline (GSK), which is teaming up with the Walter

Reed Army Institute of Research in Washington DC, and Sanofi-Aventis have both launched expanded clinical trials of vaccines in Asia and Latin America, with results expected by 2009. Trials with tens of thousands of participants would follow.

GSK's vaccine is a mixture of four live viruses that have been weakened in a process called

attenuation. The viruses should provoke a lasting immune response, but not cause disease (R. Edelman *et al.* *Am. J. Trop. Med. Hyg.* **69**, 48–60; 2003).

Sanofi-Aventis has taken a different approach, genetically engineering a chimeric virus that contains molecular bits from each strain of dengue. (F. Guirakhoo *et al.* *Vaccine* **19**, 3179–3188; 2006). **E.C.**

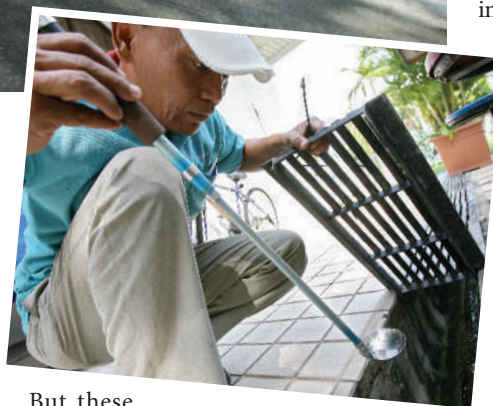
WONG MAY-EE/AP



Fever pitch: citizens of Singapore are urged to help prevent the spread of dengue fever. Officials are checking homes for signs of mosquitoes breeding.

Dengue is an almost exclusively urban disease — the WHO describes it as a “man-made problem related to human behaviour”.

“If they breed, you will bleed,” warns a pamphlet being distributed by the Singapore government that features a menacing mosquito. Unlike other countries in southeast Asia, Singapore has spent millions on mosquito control and surveillance, keeping a daily map of dengue hot spots. Citizens are fined up to S\$200 (US\$132) if mosquitoes are found on their property.



But these measures haven't stemmed the virus's spread in recent years, largely because of ineffective programmes in other countries, says Duane Gubler, a dengue expert at the University

of Hawaii in Honolulu. An influx of workers from neighbouring countries in which the disease is endemic transmits the virus to small pockets of mosquitoes in Singapore, Gubler says. Worse, decades of low infection rates have left the country's population with little immunity and now fewer mosquitoes are needed to spark an outbreak.

Until recently, there has been scant financial support for research and development of vaccines and treatments. But coinciding with an economic boom in affected cities, drug companies have begun pumping in money and manpower (see ‘Research frontline’). Dengue's immediate potential public and private market is unprecedented, says Harold Margolis, director of PDVI. In 2003, his firm received a \$55 million grant from the Bill and Melinda Gates Foundation, Seattle, Washington, to support development and testing of dengue vaccines, including gauging the costs and impact of a vaccine. And many in the field say an economic incentive improves the outlook

for a vaccine or treatment. “I'd rather be working with a company that feels positive they're going to get some reward,” Halstead says.

A dengue vaccine could have a market in the hundreds of millions, says Margolis, who is working to ensure that the vaccine will be available to everyone who needs it — not just those who can pay for it. A two-tiered market — public and private — is likely to emerge, catering to governments on one hand and to travellers, the military and foreign-aid organizations on the other, says Margolis. “You can sell it to the private sector, but it may not make dengue go away,” warns Margolis. “If that happens, that could severely affect the success of dengue vaccination.”

But a dengue vaccine is not a *fait accompli*, and a number of challenges could slow or derail development. One of the most important is the vaccine's safety. A successful vaccine would need to provide immunity to all four strains of the virus. Inadequate protection against any one strain could boost chances of coming down with DHF. And with hundreds of millions of people vaccinated, any slight safety problem will be magnified, says Halstead.

Although no one questions the need for a dengue vaccine, some wonder how far away it actually is. “We've been waiting for a vaccine for a very long time,” says Michael Nathan, dengue specialist at the WHO. He emphasizes the importance of continued mosquito control, which can address not just dengue, but other mosquito-borne diseases, such as yellow fever and malaria (see ‘Mosquito-borne diseases’).

In the meantime, Singapore and its increasingly wealthy neighbours may have to contend with more of Ng's generation falling to this indiscriminate killer.



AMERICAN CHEMICAL SOCIETY BLOG

Get our diary entries from this year's meeting in Boston
<http://blogs.nature.com/news>

HOW HWEE YOUNG/CORBIS

MOSQUITO-BORNE DISEASES

Disease	Causative agent	Genus of mosquito	Estimated annual cases	Estimated annual deaths	Estimated countries affected	Vaccine status
Dengue fever	Dengue virus	<i>Aedes</i>	50–100 million	20,000	>100	In clinical trials
Yellow fever	Yellow fever virus	<i>Aedes</i> and <i>Haemagogus</i>	200,000	30,000	>42	Available
Japanese encephalitis	Japanese encephalitis virus	<i>Culex</i>	50,000	>10,000	>10	Available
West Nile fever	West Nile virus	<i>Culex</i>	Varies from year to year, depending on outbreaks	Varies	Africa, Australia, Europe, Middle East, Asia and North America	In clinical trials
Malaria	<i>Plasmodium falciparum</i> , <i>P. vivax</i> , <i>P. malaria</i> and <i>P. ovale</i>	<i>Anopheles</i>	500 million	>1 million	>105	In clinical trials

Q&A

Geneticist trades plants for politics

Nina Fedoroff is a plant geneticist who has won many accolades for her work in crop productivity and is a staunch proponent of sowing genetically modified (GM) crops in Africa. She could now be facing her toughest challenge yet. Fedoroff is the new science and technology adviser to US Secretary of State Condoleezza Rice.

The woman who was the first to clone transposons — segments of DNA that can switch position on the genome, changing the expression of genes — is now charged with briefing Rice and other officials on the scientific aspects of foreign policy and improving science literacy in the department. But some of her well known opinions could cause friction. She's opposed to the push to produce ethanol from maize (corn), which the Bush administration supports. And her stance on genetic engineering contrasts with those of many African countries and of the Alliance for a Green Revolution in Africa, a partnership founded by the Rockefeller Foundation and the Bill & Melinda Gates Foundation.

Fedoroff, who recently won a National Medal of Science, is also an author of *Mendel in the Kitchen* (National Academy Press, 2004), which argues that GM crops are the most environmentally responsible way to feed the world.

You use the phrase 'molecular modification' rather than 'genetic engineering'. Why?

I think part of the reason that people go 'eww!' is because engineering is something that we do to buildings, not to plants. All of agriculture, all of modifying plants to suit people as opposed to their own survival, is about molecular modification.

Are GM crops high on your agenda?

Yes. I will use as much influence as I have to promote that issue. Africa was left behind by the green revolution. We are up against the limits of what we can do on the planet. The amount of land under cultivation has, to a first approximation, not changed in more than half a century. And we are adding around 70 million people a year. Where do people think the food will come from? My basic belief is that we can't solve the political problems of the world as long as we have people that can't grow as much food as they consume.

Will biotechnology help food crops adapt to climate change?

We don't know what is going to happen with climate change, but stressing plants decreases their productivity. The plant hunkers down and tries to survive. So understanding the relationship between stress and depositing sugar as starch — which is what productivity is about — is really important. How do you increase tolerance of stress, whether from

drought or extremes of temperature, without compromising productivity?

So how do you see that field developing?

It is a complicated issue and the funding is pathetic. It has now moved into the private sector, and companies are in the business of making money, not sharing science.

What's your stance on maize ethanol?

People need to understand that if you grow maize for ethanol, you drive up the price of the maize. Brazil turns sugar into ethanol and it drives up the price of sugar. Now the World Food Programme can buy less and feed fewer people. Ethanol from maize is not going to solve the world's energy problems, it is going to exacerbate them. And ethanol combustion produces the same carbon dioxide emissions as gasoline. Besides, think about the millions of years of photosynthesis that are deposited in oil that we burned through in 100 years. You can't recreate that process from an annual photosynthetic harvest.

Are you going to be able to voice your opinions publicly on matters like this without checking with the administration first?

I don't know. Ask me in six months. But I

am a person who tends to work behind the scenes.

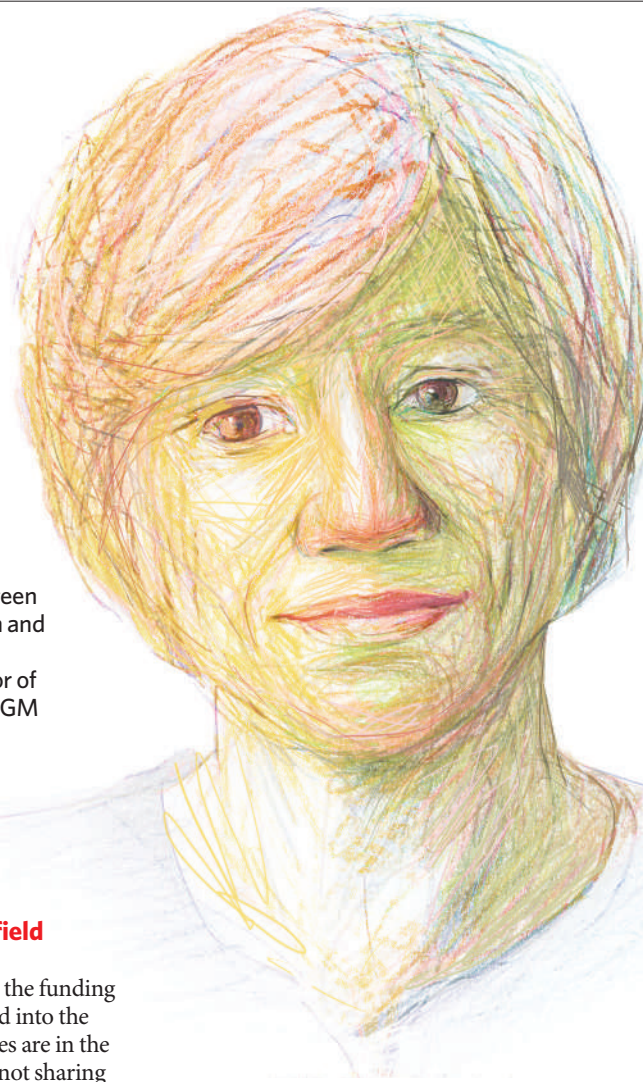
What will you focus on first?

There is a huge number of issues. On my first day of work I am going to a briefing on the state of containment of polio. There's smallpox, there's bird flu. Whether it is global climate change, agriculture, disease — all these things are so intertwined...

Do you fear that you will be asked to politicize science?

I simply don't know what I'll be asked. I think the only sensible way I can answer is that in politics short-term goals often win out over long-term goals, and that is certainly the determining factor in many decisions. I know that there are a lot of rumours about the politicization of science, and I think I would not be responsible if I said I know what the truth is.

Interview by Emma Marris



J. DEPCZYK



**RISING TEMPERATURES
'WILL STUNT
RAINFOREST GROWTH'**
Plants suffering in the heat
could make global warming
worse.
www.nature.com/news

PUNCHSTOCK
TECHNION: J. RAOUX/AP PHOTO

Achievement index climbs the ranks

A tool that ranks researchers by the quality of their work is also a good predictor of their future performance¹, according to its designer, physicist Jorge Hirsch at the University of California, San Diego.

The h-index was first proposed in 2005 by Hirsch², and attracted immense interest from scientists because of its claimed objectivity in ranking scientific achievement. Although other bibliometric measures of performance, such as counting up the number of papers published in *Nature* and *Science*, have widely acknowledged shortcomings, they are still used to assess an individual's or department's research.

Scientists generally prefer the alternative of assessment by peer review, but this process can be affected by subjectivity and bias. A fair and transparent method for gauging performance is greatly desired, especially by young scientists seeking tenure or academic placement. To this end, the British government has announced that after 2008, it will base funding assessments for universities purely on such a metric.

But which to use? Counting the number of papers might give an undue advantage to scientists who publish lots of little consequence, or in small 'salami slices'. The number of citations is a better measure of a work's impact, but it can be distorted by a few highly cited papers among many indifferent ones.

The ratio of these two measures — the mean number of citations per paper — is also a good indicator of performance³. But the h-index offers an alternative: it is the number *n* of a researcher's papers that have all received at least *n* citations. Thus, Frank Wilczek's h-index of 68 (see 'From the top down'), means that 68 of his papers have been cited at least 68 times each.

From a selection of well-cited papers in the journal *Physical Review B*, Hirsch selected 50 authors who had started publishing papers in the 1980s, and assessed how their careers had developed. He calculated how well various metrics — the h-index, the number of publications (productivity), the number of citations and the mean number of citations — during the first 12 years of their publishing career predicted the values over the next 12. But rather than assessing the cumulative impact of previous and future work, he investigated how well the indices predicted the quality of subsequent work.

Hirsch found that the h-index was slightly more accurate in this respect than the number of citations, and substantially better than the



At 110, Ed Witten (left) has a much higher h-index than Stephen Hawking's 62.

other two indices. It even predicted future productivity better than did past productivity. "Perhaps prolific authors with small citation counts feel less incentive to continue being prolific, as

they perceive that their work is not having an impact," he suggests. In other words, they decide that it's not worth trying so hard.

Hirsch also shows how the h-index can discriminate in papers with multiple authors. "I'm concerned about the blind way in which a co-author of a paper is allocated a full citation count for that paper, irrespective of his or her individual contribution," he says. The h-index automatically allocates a smaller

proportion of the credit to the researchers who were likely to have contributed less, so "it does a good job of unravelling that", he explains.

One of the h-index's benefits, Hirsch claims, is that it is hard to manipulate by self-citation — researchers would have to cite many of their own papers many times to boost their h-index rating. But physicist Michael Schreiber, of the Technical University of Chemnitz in Germany, argues that self-citations can distort the value appreciably⁴, and he proposes that they be excluded to 'sharpen' the index. James Fowler, also of the University of California, San Diego, says that even when self-citations are removed from citation counts, those who cite themselves more seem to be more successful. His group found that more than half of the citations a scientist receives can be attributed directly or indirectly to self-citation⁵.

Nonetheless, the h-index does seem to be able to identify good scientists, and it is becoming widely used informally, for example to rank applicants for research posts. "People in various disciplines are using or considering using the h-index as one of the criteria both for appointments and for grants allocation", says Hirsch. "I have seen several job applicants send us their h-index in their CV."

Schreiber agrees that "the use of the h-index is increasing, at least unofficially." Whether you like it or not, he says, "the h-index is here to stay". However, Schreiber cautions, "it is always a problem to reduce research work to one number. We should be careful using it and should be aware of its limitations." ■

Philip Ball

1. Hirsch, J. E. Preprint at www.arxiv.org/abs/0708.0646 (2007).
2. Hirsch, J. E. *Proc. Natl Acad. Sci. USA* **102**, 16569–16572 (2005).
3. Lehmann, S., Jackson, A. D. & Lautrup, B. E. *Nature* **444**, 1003–1004 (2006).
4. Schreiber, M. Preprint at www.arxiv.org/abs/0708.0120 (2007).
5. Fowler, J. H. & Aksnes, D. W. *Scientometrics* **72**, 427–437 (2007).

From the top down

Top physical scientists by h-index*.

Physics

- | | |
|--|-----|
| 1. Ed Witten | 110 |
| <i>Institute for Advanced Study, Princeton</i> | |
| 2. Marvin Cohen | 94 |
| <i>University of California, Berkeley</i> | |
| 3. Philip Anderson | 91 |
| <i>Princeton University</i> | |
| 4. Manuel Cardona | 86 |
| <i>Max Planck Institute for Solid State Research, Stuttgart, Germany</i> | |
| 5. Frank Wilczek | 68 |
| <i>Massachusetts Institute of Technology</i> | |

Chemistry

- | | |
|---|-----|
| 1. George Whitesides | 135 |
| <i>Harvard University</i> | |
| 2. Elias James Corey | 132 |
| <i>Harvard University</i> | |
| 3. Martin Karplus | 129 |
| <i>Harvard University</i> | |
| 4. Alan Heeger | 114 |
| <i>University of California, Santa Barbara</i> | |
| 5. Kurt Wüthrich | 113 |
| <i>Swiss Federal Institute of Biology, Zurich</i> | |

Computer science

- | | |
|---|----|
| 1. Hector Garcia-Molina | 70 |
| <i>Stanford University</i> | |
| 2. Deborah Estrin | 68 |
| <i>University of California, Los Angeles</i> | |
| 3. Ian Foster | 67 |
| <i>Argonne National Laboratory, Illinois</i> | |
| 4= Scott Shenker | 65 |
| <i>International Computer Science Institute, Berkeley</i> | |
| 4= Don Towsley | 65 |
| <i>University of Massachusetts, Amherst</i> | |
| 4= Jeffrey D. Ullman | 65 |
| <i>Stanford University</i> | |

*Not comparable across disciplines (multiple sources)

SNAPSHOT Trouble in paradise

Bali's first confirmed human victim of H5N1 bird flu, Ni Luh Putu Sri Windiani, is transported in a sealed coffin from Sanglah hospital in Denpasar. The 29-year-old woman from Tukadaye village in Jembrana died on 12 August. Her death brings Indonesia's total number of fatalities from bird flu to 82, and hits its main tourist destination while it is still reeling from terrorist incidents.

Windiani's five-year-old daughter Dian developed a fever after playing with chickens, and died on 3 August. Tests have not been carried out on Dian, who has already been buried. Indonesia's bird flu commission is waiting for test results from a 2-year-old girl living nearby who is also suspected of having bird flu.

In recent weeks many chickens had become sick and died suddenly in Tukadaye, and instead of burning the carcasses, the villagers fed

them to pigs or buried them, officials say. This is of special concern, because scientists think that pigs infected with both avian flu viruses and human flu viruses are a likely source of the pandemic

versions that periodically kill millions of people across the globe.

"The situation is bad," Chairul Nidom, a virologist at Airlangga University in Surabaya, Java, told *Nature*. "We don't know yet how the

infections happened. Maybe they passed through another animal."

A law preventing H5N1-infected poultry entering Bali was revoked in July by the Home Ministry.

David Cyranoski



F. LISNAWATI/AP

Transparency urged over research payments

Charles Grassley, the Republican senator for Iowa, has said that he will introduce legislation to set up a national public registry of payments made to doctors by pharmaceutical companies.

Four states and the District of Columbia already have laws requiring such disclosures. But, Grassley says, such laws should be rolled out nationwide. "For the sake of transparency and accountability, shouldn't the American public know who their doctor [*sic*] is taking money from?" Grassley asked in a speech on the Senate floor on 2 August.

Grassley is not the first legislator to propose a registry, and his office did not provide details of his plan. Whatever the plan entails, it is likely to meet resistance from the American Medical Association and the Pharmaceutical Research and Manufacturers of America.

Public unease with lucrative arrangements

between researchers and industry has grown in recent years, increasing demand for more information about these ties. Questions about drug safety have highlighted the issue of whether industry-sponsored research by academic scientists is trustworthy. And a 2003 scandal over conflicts of interest at the National Institutes of Health, Bethesda, Maryland, embarrassed the agency and led to predictions that the practice would spill over to universities (see *Nature* **431**, 725; 2004).

Now, this seems to be occurring. In his speech, Grassley acknowledged that universities often require their researchers to disclose payments from outside sources, but said there are flaws with this system. For instance, it relies on the researcher to make accurate disclosures, which are not generally released to the public. Now, Grassley is asking universities for more information about how well the process works.

Observers say a national registry should improve on the shortcomings of existing state registries. "The states deserve credit for being a source of innovation on this issue, but not one of them has produced complete or readily accessible disclosure," says Peter Lurie of the watchdog group Public Citizen, in Washington DC.

An analysis of drug payment disclosures in Vermont and Minnesota has found serious loopholes and a lack of enforcement (J. S. Ross *et al.* *J. Am. Med. Assoc.* **297**, 1216–1223; 2007). For instance, 61% of payments to Vermont physicians were not made public because the state's law allows companies to designate the information as a "trade secret". And only one-quarter of drug companies in Minnesota reported their payments, which did nothing to ensure compliance with the law, Lurie says. ■

Erika Check

Stem-cell institute's hunt for president goes on

California's \$3-billion stem-cell initiative has given up on its bid to find a permanent president by this summer.

On 8 August, the board at the California Institute for Regenerative Medicine (CIRM) appointed Richard Murphy as the agency's interim president. Murphy, former president of the Salk Institute for Biological Studies in La Jolla, California, has begun working on a six-month contract but could stay until March 2008.

Murphy was a member of the CIRM's board until he retired from the Salk in July. He did not want to be considered for a permanent post at the CIRM because he is relocating to the US east coast.

His temporary appointment has not been unanimously welcomed by the CIRM board. Some fear it will distract the agency from its urgent need to find a replacement for former president Zach Hall, who resigned in April. The agency has already begun handing out its first rounds of research grants (see *Nature* 446, 238–239; 2007).

South African president fires deputy health minister

Nozizwe Madlala-Routledge, South Africa's deputy health minister, was sacked last week by President Thabo Mbeki ostensibly because she travelled to an AIDS vaccine conference in Spain without receiving the required permission to make the trip.

Her dismissal has been interpreted by opposition members and AIDS activists as a deliberate snub of her forthright views on AIDS (see page 727). Madlala-Routledge is the only minister or deputy minister to have publicly taken an AIDS test.

Earlier this year, Madlala-Routledge took the lead in promoting AIDS awareness in South Africa while her boss, health minister Manto Tshabalala-Msimang, was ill. The two have reportedly clashed in recent months over the government's AIDS policy. Tshabalala-Msimang has promoted use of garlic and other nutritional products to fight AIDS.



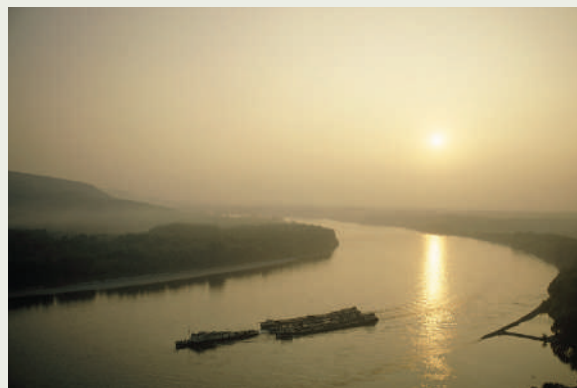
Sacked: Nozizwe Madlala-Routledge.

Survey team sets sail for a summer on the Danube

The world's biggest river research expedition for 2007 was set to launch this week on the Danube (pictured). An international team including 18 scientists will travel 2,375 kilometres on a flotilla of three boats from Regensburg, Germany, down to the river's delta in Romania and Ukraine.

In 1994, countries through which the Danube flows signed an agreement to work together to improve water quality.

The signatories are funding the project, which has a budget of more than €1 million (US\$1.4 million), as a way to provide highly reliable and comparable data across the various borders.



J. P. BLAIR/NATIONAL GEOGRAPHIC/GETTY

Novartis loses claim to patent Gleevec in India

In a closely watched ruling for the pharmaceutical industry, a high court in Chennai, India, has dismissed a patent claim for the anticancer drug Gleevec (imatinib).

The Swiss drug firm Novartis was seeking a patent for a newer version of the drug. It had gone to court after the patent office in Chennai rejected its application last year on the grounds that Indian law did not allow patenting of incremental modifications of existing compounds. Novartis argued that the law does not conform with the Trade-Related Aspects of Intellectual Property Rights (TRIPS) agreement.

Granting the patent, say opponents, would have set a legal precedent discouraging firms from making cheaper, generic versions of life-saving drugs.

Ranjit Shahani, managing director of Novartis India, has warned that the verdict is a loss for India. He adds that multinational drug companies will instead invest in research in China.

Freed medics were forced to confess in Libya

The six medical workers freed last month after eight years in a Libyan jail were tortured while in detention, a son of Libyan leader Muammar al-Gaddafi has admitted.

In an interview last week with the television network Al Jazeera, Seif al-Islam al-Gaddafi confirmed allegations that the medics — a Palestinian-born doctor and five Bulgarian nurses — were forced to confess that they deliberately infected more than 400 Libyan children with HIV. "They were tortured by electricity and they were threatened that their family members would be targeted," he said. The Libyan government has since issued an official denial.

Libya's highest court last month commuted death sentences against the six to life imprisonment after families of the children each received US\$1 million in compensation. This opened the way for the medics' release on 24 July (see *Nature* 448, 398; 2007).

Seif al-Islam conceded that investigations were not carried out in a professional way, but denied that Libya would face legal action. Ashraf Ahmad Jum'a, the freed doctor, plans to file a complaint against Libya before a United Nations human-rights panel, his lawyer said last week.

Logo offers physicists a system for hire education

Looking for a job in applied physics? Then say so when you give your next talk. That's the idea behind an initiative from the Japan Society of Applied Physics, which is encouraging postdocs and graduate students to advertise their availability on their PowerPoint slides.

Thanks to programmes in the 1990s, Japan now has a glut of more than 15,000 postdocs — but too few employment opportunities for them (see *Nature* 447, 1028; 2007).

So the physics society has come up with a 'Career Explorer' logo, which it recommends postdocs incorporate on the opening slide of presentations on their research. The first opportunity to test the logo, which features a smiling young researcher peering through enormous binoculars, will come next month at the society's annual meeting in Sapporo.

Correction

The News story 'Academic accused of living on borrowed lines' (*Nature* 448, 632–633; 2007) incorrectly identified Peter Weingart as a co-worker of Hans-Werner Gottinger. Weingart did not work on any research projects with Gottinger, but was simply another member of the department of sociology at the University of Bielefeld at the time.

BUSINESS

Powerful incentives

Can an expanded loan guarantee programme dispel US power companies' hesitation about resuming construction of nuclear power plants? **Geoff Brumfiel** investigates.

To an outside visitor, America's nuclear power plants look distinctly dated. Operators watch their reactors' progress on racks of indicator lights and run them in part by using old-fashioned relay switches. Even the beige and taupe colour scheme of the control rooms is a throwback to an earlier era.

The vintage look underscores the screeching halt that nuclear-power business came to in 1979, when a successfully contained accident at Three Mile Island in Pennsylvania soured the public's attitude to nuclear energy. Since then, work hasn't started on a single new nuclear power station in the United States. But as memories of that era fade and concerns mount about global warming, the industry thinks that attitudes are changing.

In an effort to restart construction, congressional supporters of the industry such as Senator Pete Domenici (Republican, New Mexico), have incorporated powerful incentives into a pending energy bill. If the bill passes with these provisions intact, as is considered likely, it would allow the government to provide tens of billions of dollars in federal loan guarantees to the industry. That would allow utilities to raise the money needed to build new plants, says Mitchell Singer, a spokesman for the Nuclear Energy Institute (NEI), an industry lobby group in Washington DC.

Risky guarantees

Under the legislation, the federal government could be liable to pay back loans covering up to 80% of construction costs if the utility defaults. Not everyone thinks that this is the best course of action. "This is a huge risk for taxpayers," says Michele Boyd, legislative director at Public Citizen, a consumer advocacy group in Washington DC.

For investors, new nuclear plants offer an unfortunate combination: they're both long term and high risk. The NEI estimates the cost of a new facility at US\$3.5 billion to \$5 billion. For a typical utility with a market capitalization of, say, \$50 billion, that's a big investment to make.

Lenders and investors are reluctant to finance such projects because of the industry's



Three Mile Island, 1979: the Pennsylvania accident brought construction of nuclear reactors to a halt in the United States.

troubled past. In the 1960s and 1970s, construction of nuclear plants regularly ran behind schedule and over budget. As inflation rose in the 1970s, and energy demand subsequently shrank, utilities struggled to make payments on their plants. "There were significant write-offs and massive dividend cuts," recalls Caren Byrd, an energy analyst at Morgan Stanley in New York. "The financial community was hurt very badly."

The Bush administration has already offered a swathe of subsidies to encourage new nuclear plants. Since 2004, the US Department of Energy has spent \$250 million on collaborations with industry to certify a new generation of reactor designs and prepare the first licence applications. A 2005 energy bill offers utilities a production tax credit of 1.8 cents for every kilowatt-hour — around one-quarter of the typical electricity price — generated by the first 6,000 megawatts of nuclear generating capacity for the first eight years of operation. It also offers them insurance of up to \$2 billion for the

first six plants in the event of construction and licensing delays.

These incentives have led four companies to file for 'early site permits'. But only one firm — Constellation Energy of Baltimore, Maryland — has started to apply for the 'combined operating licence' needed to start construction. Funds to start construction have yet to materialize, says David Schlissel, an analyst at Synapse Energy Economics, an energy consultancy firm based in Cambridge, Massachusetts. "It's unclear that anyone will put up the capital for the new plants unless the Feds guarantee the loans," he says.

The NEI claims that the industry would be ready to borrow about \$50 billion to build new plants within two years if such guarantees were available. It also says that utilities would protect the government's investment by paying into a fund to cover losses if a couple of projects fall through. The entire system "will be self-financing in many ways", Singer claims.

But a study in June by the non-partisan Congressional Budget Office in Washington DC warned that new nuclear plants pose "significant technical and market risks". It also said that the US Department of Energy, which would oversee the programme, was likely to underestimate the fees that need to be paid into the fund.

Safety net

The guarantees would provide a major boost for plant construction, says Marilyn Kray, vice-president for project development at Exelon, a utility based in Chicago, Illinois, and the largest nuclear generator in the nation. They would reassure lenders, and allow utilities to borrow at lower rates. Given the enormous capital costs, he says, "a single interest percentage point is quite significant."

"It would be a very useful incentive to have," agrees Dimitri Nikas, an energy analyst with Standard & Poor's, a financial services company in New York. But it might still fail to drive down the costs of construction to a competitive level. The expert labour and technology needed to build such plants is expensive, as is the meticulous regulatory process. The bottom line, Nikas says, is that the incentives may get one or two plants built — but they won't herald a building boom in nuclear power stations. ■

W. MCNAMEE/CORBIS

SECRETS OF THE MARTIAN SOIL

For 30 years scientists have believed that there are no organic molecules in the martian soil. Will NASA's Phoenix probe prove them right or wrong, asks **Corinna Wu**.

In the summer of 1976, humanity first got its surrogate hands on the soil of another planet. Two Viking probes scooped up samples from their landing sites and fed them into an array of instruments carried from Earth for the purpose. Perhaps the most technically sophisticated of those instruments were the gas chromatograph-mass spectrometers (GC-MS) designed to detect evidence of organic compounds in the soil. After months of operation they had found none at all. The Viking missions were accepted almost unanimously as showing that Mars was a sterile planet, and the GC-MS data were a crucial part of the evidence.

This month, another mass spectrometer is on its way to Mars. NASA's Phoenix mission was launched on 4 August, and if all goes according to plan, it will touch down 1,200 kilometres from the north pole of Mars on 28 May 2008. Phoenix is essentially a sister ship to the Mars Polar Lander (MPL) lost in December 1999 — hence the imagery of its name — and the first in NASA's new line of low-cost 'Scout' missions. It is a Buick to the Bentleys of the Viking missions, but some chemists are all but certain that it will find evidence for organic chemicals close to the martian surface even though the better equipped Vikings found none. If it does it will raise the hopes of those who cling tenaciously to the idea that there might be evidence for past, or even present, life somewhere on that frigid, radiation-battered and deeply inhospitable planet.

The Viking landers each carried a suite of three experiments designed to detect micro-organisms in the dust and sand that, along with pumelled rocks, made up the surface at the landing sites. Their results were inconclusive. When watered and fed nutrients, the martian soil gave off gases in ways that had not been expected, some of which might be consistent with biological activity. But the GC-MS data were unequivocal.

The GC-MS is a workhorse lab instrument



used to pick out the various components in mixtures of organic molecules. It works by first passing samples through a thin capillary column — the chromatograph. Small molecules move through quickly; heavier ones more slowly. When the compounds emerge from the column, sorted by size, they pass through the mass spectrometer, which measures the mass either of whole molecules or their fragments.

The Viking GC-MS team was led by Klaus Biemann, a chemistry professor at Massachusetts Institute of Technology (MIT), Cambridge, and a pioneer in the development of such instruments. Biemann was not a space scientist; his work concentrated on determining the structure of protein fragments — laying the foundations for today's 'proteomics'. He agreed to work on the Viking instrument only as an act of "scientific charity", he recalls. "I said if it needs to be done, it might as well be done well." And it was.

At the time, the GC-MS Biemann had in his laboratory was the size of a room. The instrument needed to analyse the compounds that might be given off when martian soil samples

were heated had to fit into a box just thirty centimetres long on each side. Biemann's team produced a fully automated system that met the constraints in size and available power and yet was still sensitive to compounds present only at a parts-per-billion level. "To get two machines a hundred million miles out and both of them working is just a marvel of engineering," says Steven Benner of the Foundation for Applied Molecular Evolution in Gainesville, Florida, who has worked on chemical approaches to various astrobiological questions.

Organic ball game

Biemann's team expected to have organic compounds to analyse when its instrument reached Mars. Mars's cratered face showed it had been bombarded by meteorites, and by the 1960s some meteorites were known to contain organic compounds. So organic compounds were to be expected, even if Mars itself had produced none of them. "You should be sitting in a sea of this stuff," says Benner. Indeed, according to Benner some Viking developers worried that the GC-MS would be swamped by the sheer amount of organics.

Yet the only things that came out of the heated soil samples on Mars were water and carbon dioxide, which were taken to have been either physically trapped in soil particles or released from inorganic minerals. Although the experiments did see some organic compounds, they were those that had been used to clean the equipment back on Earth, and had already been detected when the instrument was tested in deep space during the probes' transit to Mars.

With no evidence for organic molecules in the soil, the results from the life detection

experiments were to some extent rendered moot. “That’s the ball game,” said Jerry Soffen, Viking’s project scientist, at the time. “No organics on Mars, no life on Mars.” The activity seen in the samples was subsequently interpreted as being due to chemistry, not biology. Bombardment by ultraviolet light would make the soil rich in oxidants such as peroxides, and also lead to some unusual — on Earth — compounds such as carbon suboxide (C_3O_2); scientists proposed various ways that reactions with such chemicals could have provided the results seen by the life detection experiments. Only one of the investigators, Gil Levin, continued to think that what they had seen was best explained by biology.

In 2000, though, Benner, then at the University of Florida, Gainesville, suggested that Mars might indeed have contained some organic material, but that Viking could have missed it. He proposed that the organics might take the form of mellitic acid, which is seen when the mess of polymerized carbon that makes up much of the organic component in meteorites is not completely oxidized. “It is readily formed under the oxidizing conditions you’d expect to find on Mars,” Benner says. “It is quite stable to further oxidation, and it’s also refractory — it doesn’t dissolve in anything. And when you heat it, it doesn’t give off anything volatile. So they could have been sitting in a sea of this stuff and not seen it.” Benner’s study estimated that this process could have generated 2 kilograms of mellitic acid per square metre of martian surface over the course of 3 billion years¹. Unfortunately, it takes a lot of heat to break down mellitic acid — and when it breaks down, the primary product is benzene, says Benner, which was one of the solvents that had been used to clean the instrument. If the Viking experiments had heated their samples to 600 °C instead of 500 °C, they might have picked up traces of something distinctive. But they didn’t reach 600 °C.

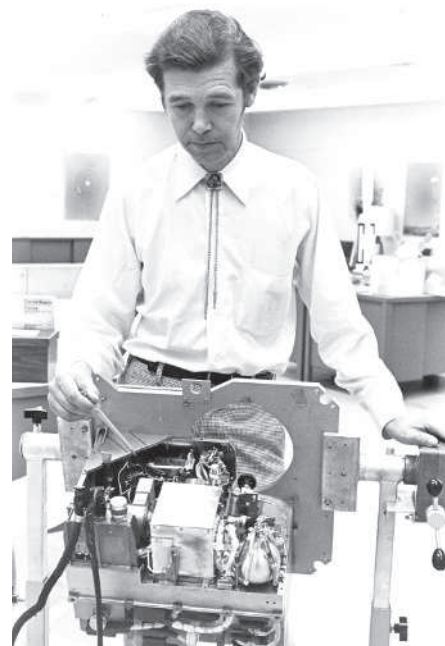
“That’s the ball game. No organics on Mars, no life on Mars.”
— Jerry Soffen

In October 2006, a team including chemist Rafael Navarro-González of the National Autonomous University of Mexico’s Institute of Nuclear Science, Mexico City, and planetary scientist Christopher McKay of NASA’s Ames Research Center in Mountain View, California, reported that the experiment might not have picked up some other types of organic compound in the soil².

Navarro-González was inspired to revisit the Viking tests for organic molecules after the Opportunity rover discovered jarosite, a hydrated iron sulphate that forms in the presence of water, on Mars in 2004. Studying jarosite-containing soils in the Rio Tinto area of Spain, he found that getting organic material out with chemical approaches was relatively easy — but getting it out just by heating was not. “When I repeated the Viking experiments,” he says, “I was surprised to see that despite the huge amount of organic matter present, there was virtually no detection of organics in the sediments. This was quite strange.”

Independently, McKay had been doing research on soils from the Atacama desert in Chile, and had also started to suspect that the Viking experiments weren’t telling the whole story. Alison Skelley, a graduate student at the University of California, Berkeley, had asked McKay to review a paper on a device she had developed for detecting amino acids in soil³. McKay found the paper striking, noting that “it found that there were a thousand times more amino acids released by chemical extraction than pyrolysis” — the heating method used by the Viking experiments. Then McKay says, “Within a month, Rafael told me about his puzzling result with the jarosite. That’s when I suggested that we ought to see if this effect was widespread.”

In addition to the Rio Tinto sediments and Atacama desert samples, they tested soil samples from other inhospitable and vaguely martian environments — the Dry Valleys of



H.-P. BIEMANN

Klaus Biemann with a copy of the Viking GC-MS in 1975, the day before the mission landed.

Antarctica and the Libyan desert. Chemical extractions revealed low levels of organic compounds — between 20 and 90 micrograms of carbon per gram of soil, in Antarctica for example. But heating samples from most of the sites to 500 °C did not produce organic volatiles that their GC-MS setup — a combination of several commercially available instruments — could detect. Only at 750 °C did they start to get a signal from more than half the samples — a temperature the Viking systems were not designed to reach.

Heated debate

Navarro-González and McKay think that during the heating step of the Viking experiment, any organics given off at moderate temperatures would have been turned into CO_2 before they reached the GC-MS, thanks to catalytic iron compounds in the martian soil. “We suggest that a small portion [of the carbon dioxide seen by the Viking experiments] could have resulted from the oxidation of organics,” says Navarro-González. “Even if it’s just a small percentage, this could mean levels of organics on the surface of Mars a thousand times higher than expected.”

Before publication they sent the paper to Biemann for his feedback and “he was completely upset,” says Navarro-González. “But he did bring up some interesting points that we had to resolve.” Navarro-González says they sent further versions of parts of the paper; Biemann says he received no satisfactory answer to his questions. In the finished article, the authors thank Biemann “for comments on an earlier version of the manuscript”.

Biemann, who denies he was upset, says he heard that the Navarro-González paper had finally been published only after his daughter read a story about it in the *Wall Street Journal*. Being thanked in the paper, he feels, implies that



Spot the difference: the Atacama Desert in Chile (left) shares many features with the surface of Mars.

he agreed with their second version, which he says he never saw. And so in a strongly worded critique of the revisionist work, Biemann argued that the experimental setup used by the Navarro-González and McKay team was a thousand times less sensitive than the Viking device⁴. “To say in their experiment that they don’t find things at 500 °C that they find at 750 °C doesn’t mean anything,” he says. If their instrumentation had corresponded more closely in its performance to the carefully-tailored Viking GC-MS, it would have been easily sensitive enough to detect what was going on at 500 °C, he says.

Biemann thinks that a misplaced zeal to find life on Mars has driven scientists, including the Navarro-González team, to try and prove the GC-MS results wrong; they want to “get rid of that obstacle,” he says. McKay, for his part, expresses frustration over what he feels is a misunderstanding of the thrust of their paper. “If I were rewriting the paper, I would emphasize that the GC-MS operated flawlessly. The problem is the pyrolysis release of organics.” He says the debate over the Viking results has been so narrowly focused on the GC-MS that the pyrolysis step has been ignored.

Even accepting the group’s analysis that there might be some organic matter in the soil, “it’s not a rich soil by any stretch of the imagination,” says McKay. “When we say [the organic fraction] could be as high as a part per million, it’s important to note that it could also still be zero.” What’s important, he says, is to take the new work into account on future missions.

Indeed, the scientists in charge of the instruments on NASA’s Phoenix lander have paid heed. The primary goal of the mission is to characterize ice and minerals in the martian soil, but the lander will also have the ability to detect organics. The Thermal and Evolved Gas Analyzer (TEGA), originally developed for MPL, heats soil samples at a constant rate, measuring changes in the rate of warming so as to detect phase changes — when things melt or evaporate they can absorb heat without changing temperature. But Phoenix’s TEGA, unlike that on MPL, also boasts a small mass spectrometer which will be used on the output from samples heated as high as 1,000 °C — twice the temperature of Viking’s ovens, and hot enough to decompose the most refractory compounds.

“If we don’t see organic compounds, we will have at least answered the question that it’s not because they are of a refractory nature,” says William Boynton of the University of Arizona in Tucson, who is lead scientist for the TEGA



Take off: Phoenix begins its journey to Mars.

project. “It will also mean that this particular environment was not suitable for protecting organic molecules from being destroyed.”

The search for life continues

Benner is more or less convinced that the Phoenix ovens will find the mellitic acid and associated salts he has predicted — which would make it third time lucky for him. He originally hoped to see evidence for the compounds from Raman spectrometers that would have been carried on the Spirit and Opportunity rovers, but a tight schedule saw the necessary lasers cut from the payload. The Beagle 2 mission should also have measured organics in the soil, and had a cunningly contrived device that would have let it take samples from underneath rocks, where the ultraviolet-induced oxidation might not be so bad. But contact with the spacecraft was never established after it left its mothership, the European Space Agency’s (ESA) Mars Express.

If TEGA does detect organics that reveal themselves only at high temperatures, the chances are strong that they will be from meteorites, not native to Mars. But their persistence would show that the soil was not quite as powerful an oxidant as the post-Viking consensus supposed, which might offer hope that in some places native organic matter might be preserved. What’s more, there is a possibility that the icy soil on which Phoenix is hoping to land might be such a place. If the source of the ice — the presence of which was confirmed by the gamma-ray spectrometer that Boynton flew on NASA’s Mars Odyssey orbiter — is ground water that has welled up and then frozen, it might contain organics derived from reservoirs that are below the reach of the soil’s harsh oxidizing properties and subsequently

protected by the ice, Benner suggests. If such organics were detected it would not necessarily prove the Viking results wrong — it would just show that different environments in different parts of Mars offer different levels of comfort for such molecules.

Navarro-González and McKay are both working on the Sample Analysis at Mars (SAM) package that will be part of NASA’s Mars Science Laboratory, scheduled for launch in 2009. SAM will contain the first GC-MS since Viking — Phoenix has no chromatograph — as well as a laser spectrometer, and it will use chemical extraction as well as pyrolysis, allowing the two techniques to be compared and to be used in a complementary way.

In the longer run, some scientists are wondering if there are robust ways to tell whether any organics found are the product of living beings, but they recognize it will be a hard problem. On Earth, it is possible to distinguish life’s organic products by the ratio of different carbon isotopes they contain, and SAM should be able to do this — Phoenix might, too.

But scientists do not expect that interpretation of any such results from Mars would be straightforward, since non-biological processes can also have isotopic signatures. An instrument on the ESA’s ExoMars mission, slated for launch in 2013, will have the capacity to measure the ‘handedness’ of any organic molecules it finds. On Earth, life uses left-handed amino acids and right-handed sugars, and samples that reflected a similar prejudice might be seen as good evidence of a living source. But if the molecules have been around for millions of years, they may well have spontaneously rearranged themselves into a random mess, and thus become indistinguishable from molecules created through non-biological processes.

Before that next level of uncertainty and debate can be reached, though, evidence is needed that there are organic molecules out there to study. If Phoenix, unlike its ill-fated sibling, survives and sends back data, it will at least have moved the debate on. “It’s more important to be looking forward to future missions than to be stuck on a debate about Viking,” McKay says. “If we get results from Phoenix and from SAM, then people will have something new to argue about.” Benner, meanwhile, just looks forward to what he sees as the inevitable surprises: “Every time we go back, it’s like going for the first time.” ■

Corinna Wu is a freelance writer in Washington DC.

1. Benner, S. A., Devine, K. G., Matveeva, L. N. & Powell, D. H. *Proc. Natl. Acad. Sci. USA* **97**, 2425–2430 (2000).
2. Navarro-González, R. et al. *Proc. Natl. Acad. Sci. USA* **103**, 16089–16094 (2006).
3. Skelley, A. M. et al. *Proc. Natl. Acad. Sci. USA* **102**, 1041–1046 (2005).
4. Biemann, K. *Proc. Natl. Acad. Sci. USA* **104**, 10310–10313 (2007).

“This could mean levels of organics on the surface of Mars a thousand times higher than expected.” — Rafael Navarro-González

SPACE INVADERS

Space exploration usually means leaving Earth's orbit. But chemists are now burrowing inside solids to open new vistas. **Katharine Sanderson** reports from the internal frontier.

Solids have a reputation for permanence and reliability. Astronauts and sailors rejoice in returning to solid ground. And few things are more durable than solid gold. Unlike their flightier cousins, the gases and liquids, solids have regular structures that generally resist deformation. But there are materials that challenge our notions of a solid. They are rigid and strong, yet flexible, incredibly light yet so porous that the surface area of one gram could cover an entire football pitch.

These solids are known as metal-organic frameworks, or MOFs. They have no internal walls, just a bare molecular scaffold, creating a regular, sturdy, crystalline structure that is packed full with empty space. But if the scientists building these structures get their way, the materials will soon be packed full of useful gases, such as compressed hydrogen or methane, making it easier to store and transport these alternative fuels.

Chemists expected that the vast, open structures of MOFs — more air than solid — would collapse readily. Yet the structures can be designed such that they are held together entirely by strong chemical bonds — carbon-carbon, carbon-oxygen or metal-oxygen. They are composed entirely of joints, made from metal oxide units, and struts, which are organic groups. As long as those metal oxide joints don't shear or buckle under pressure, the structure stands firm, like builders' scaffolding.

Structural gymnastics

The resulting wall-less structures produce solids that are incredibly light with very low densities — sometimes as low as 0.2 gram per cubic centimetre — allowing a chunk of the material to float on water. Some cleverly engineered MOFs can even flex and bend slightly without collapsing.

With all that space to play with, these solids

are more easily penetrated than most. Chemists have used porous materials, including clays and zeolites, to trap and store gases for decades. But the pores and channels in naturally occurring porous materials are of varying sizes, so researchers have sought to make porous structures with uniform openings. In trying to construct such materials in the 1990s, Omar Yaghi at the University of California, Los Angeles, hit on a recipe for making frameworks with precisely controlled pore sizes. In 1998, Yaghi engineered the first MOF structure by mixing together two molecular building blocks — namely metal oxide and organic groups¹.

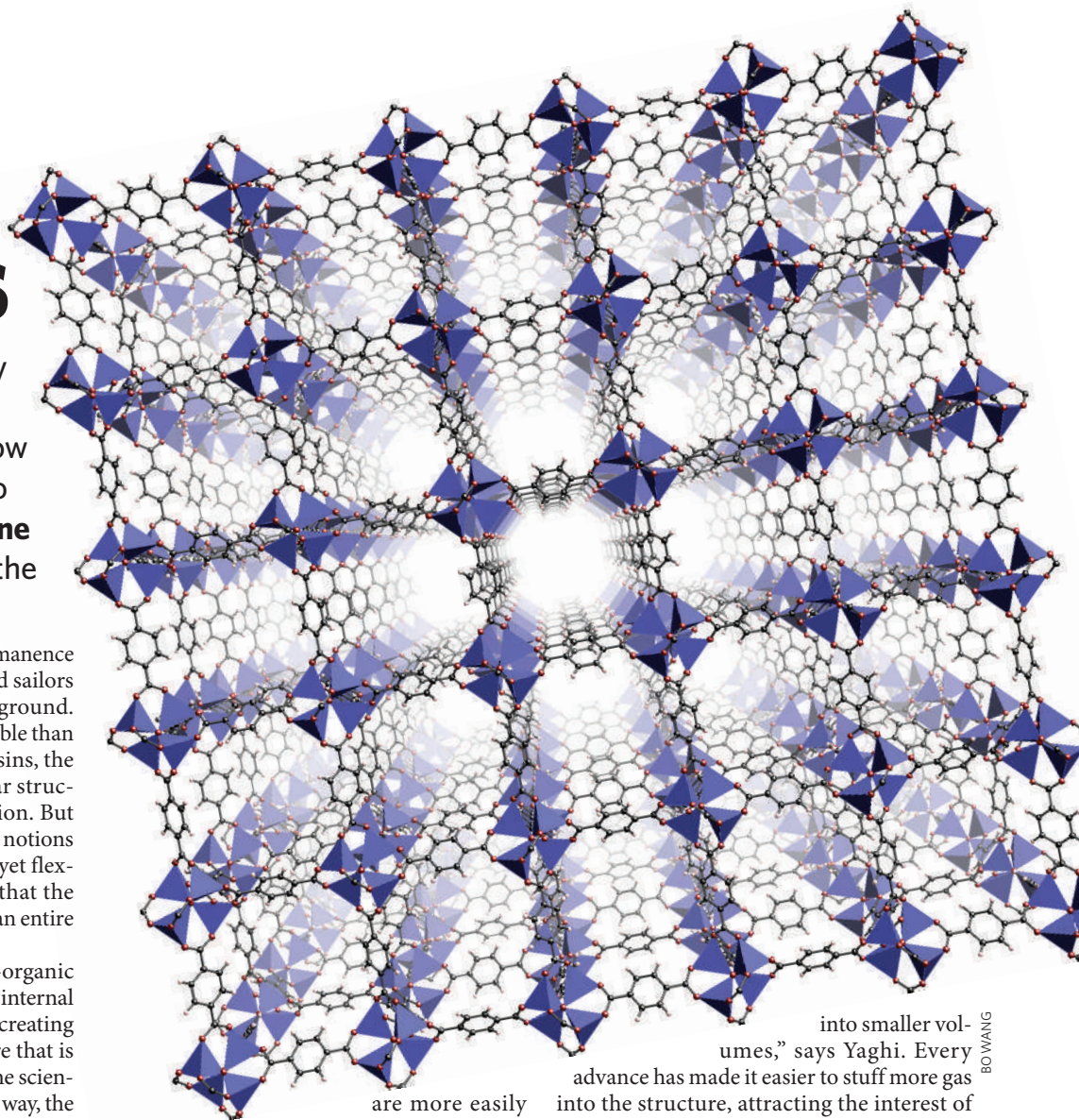
Since then, researchers have created materials with larger and larger pores. In doing so, they have repeatedly broken records for internal surface area in solids, and for low density, making zeolites look stodgy by comparison. But record-busting is not the only motivation to get bigger surface areas. "It's not just an obsession of mine to get higher numbers, it's the way to compact more and more gases

into smaller volumes," says Yaghi. Every advance has made it easier to stuff more gas into the structure, attracting the interest of German chemicals giant BASF, which is about to move its MOF research into small-scale production.

Yaghi is thrilled by BASF's investment in his work. In Yaghi's lab, MOFs can now be constructed to order. He has compiled a list of metal fragments to act as the scaffold joints, and a compatible range of organic links. If these linking groups are linear — that is with connecting points on both ends — they can be used to make chains or cube-like structures. More exotic networks can be made from triangular or square groups with connectors sticking out from each corner.

BASF supports another MOF chemist, Gérard Férey at the Lavoisier Institute in Versailles, France. Férey claims to hold the world record for the internal surface area of a MOF², published in 2005, at 5,900 square metres — or, the same area as an average-sized football field — for every gram of material. His material was built from triangular groups of chromium atoms linked together by terephthalic acid molecules.

"The limits are in our imagination."
— Gérard Férey



BO WANG

Yaghi could soon beat Férey's record. His latest MOF — MOF-200 — looks likely to have a whopping surface area of 8,000 square metres per gram, once the synthesis is complete. But Mark Thomas, an expert in porous solids at the Newcastle University, UK, cautions against reading too much into these claims, which are based on calculations developed for materials with much smaller pores, and with internal walls. "It's a pore volume converted into a surface area," he explains, "it is an apparent surface area." Férey agrees. "The numbers don't represent anything physically," he says. "We have to find a new definition for the surface of very large pores."

Still, Yaghi says that the numbers do have meaning, and they can be used to compare MOFs with each other and with other porous materials. The surface area of a single layer of porous carbon, when calculated in a similar way, comes out at a relatively tiny 2,600 square metres per gram.

Storage space

Yaghi's hopes for trapping and storing gases within the cavernous space provided by MOFs are focused on hydrogen storage. Hydrogen offers a low-polluting alternative to petrol as a fuel for road vehicles, but it faces many technical and infrastructure barriers. Not least of these is storing enough hydrogen on a vehicle in a safe and affordable way.

The US Department of Energy says the ultimate hydrogen car will need enough fuel for a 480-kilometre trip, and yet the fuel tank must not be too heavy or too bulky. It has set a target of 2010 for a viable storage system in which hydrogen makes up 6% of the system's total weight. In addition, the system has to operate over a range of "expected ambient conditions" — storing and releasing hydrogen at temperatures between -30°C and 50°C and at a maximum onboard operating pressure of 100 bar.

None of the existing options achieve this. Conventionally, hydrogen has been stored as a gas by keeping it in a high-pressure canister, or stored as a liquid at a chilly 20 kelvin (-253°C). But high-pressure tanks are bulky, and liquid hydrogen is expensive to keep cool. In principle, porous materials can increase the amount of hydrogen stored in a given volume, without relying on extremely high pressures or low temperatures.

In the late 1990s, carbon nanotubes generated much excitement as a way to store hydrogen. Researchers in industry and academia claimed storage levels ranging from 3% to 10% by weight at ambient temperatures and pressures. But most of these results — including one claim that graphite fibres can adsorb their own weight in hydrogen — have not been repro-



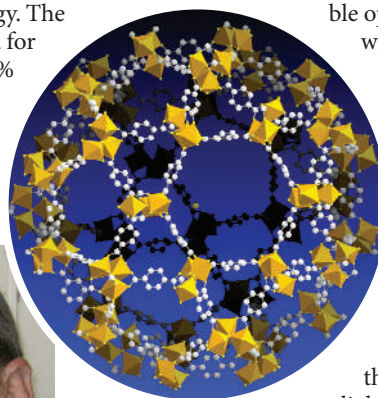
Chemical creations: Omar Yaghi was the first to design a metal-organic framework.

duced. Researchers are now more cautious of spectacular claims about hydrogen storage.

Today's front-runners for hydrogen storage in materials are metal hydride systems. These operate on the principle of chemisorption — whereby hydrogen is trapped by forming a chemical bond with the material, so forming a metal hydride. They operate at moderate pressures, but the material needs to be heated to release the hydrogen, thereby consuming more energy. The current storage record for metal hydrides is 9% by the weight of the material alone, but the Department of Energy's target includes the fuel tank



Inside story: Gérard Férey's metal-organic framework (right) is credited as having the largest internal surface area.



and accessory parts in the calculations, so that value translates into just 4% by weight for the overall storage system.

A different process, known as physisorption, is responsible for the weak electrostatic attractions that hold gases inside MOF pores. Weak interactions mean that releasing the gas from the material requires less energy — nothing like that needed to break chemical bonds. "The good thing about physisorption is that it's easily reversible," explains chemist Matthew Rosseinsky from the University of Liverpool, UK. And the great thing about MOFs that enhances that easy reversibility is their openness. "Things move in and out with great facility," says Yaghi.

For now, though, trapping hydrogen in MOFs still requires low temperatures. Although MOFs need pressures of only 70 bar to get hydrogen into the materials, they typically need temperatures around that of liquid nitrogen, at 77 kelvin — to keep them there. This is because of the weak interaction between hydrogen and the scaffold. As it starts to warm up, the energy in the system breaks the interactions, leaving the hydrogen free to exit the framework.

Out cold

Yaghi claims the record for hydrogen storage in MOFs. He says that at 77 kelvin, and at a pressure of 70 bar, his MOF-177 material can store 7.5% hydrogen by weight³. Although that sounds promising, carrying tanks of liquid nitrogen as well as hydrogen is not a viable option. "That's not very practical when you want to run an automobile," Yaghi admits. And even at these temperatures, MOF-177 still misses the 2010 target — which sets a goal for the material alone of 9% by weight.

Yaghi is optimistic about achieving the 2010 target with a MOF that can operate at room temperature. Férey is more sceptical. He thinks that the first step should be to move to slightly warmer temperatures: "If we reached 150 kelvin instead of 77 kelvin it would be a great improvement."

And recent work has challenged the idea that bigger pores are always better. Martin Schröder at the University of Nottingham, UK, has calculated the optimum pore size for stuffing hydrogen into a MOF — with the surprising result that medium-sized, rather than giant, pores were the winners⁴. With smaller pores, the thinking goes, the scaffold is more curved, improving its chances of interacting with gas molecules.

N. K. YAGHI

SCIENCE
C. ROSSLAND

It is the lure of engineering MOF structures to enhance these interactions that attracts Yaghi. If MOF structures could be tweaked to strengthen physisorption, they might be able to operate at higher temperatures. But this is a delicate art: if physisorption is too weak, the hydrogen won't stick, yet as it increases more energy is needed to release the gas. What's needed is something in between a physical electrostatic interaction and a chemical bond.

Weighty advantage

Yaghi proposes modifying the organic groups in his MOFs by adding a light metal such as lithium. The extra electrons provided by lithium would strengthen the interactions between the gas and the MOF. Yaghi predicts that by simply adding some lithium into his system he could improve the current room-temperature record of 1% hydrogen by material weight to 4%.

Others are exploiting MOF's ability to flex and bend to create clever hydrogen-trapping devices. Thomas and Rosseinsky and their colleagues constructed a cylindrical MOF from a network of large cavities separated by small windows⁵. The windows are too small to let hydrogen through. But when the structure is completely dried out under vacuum, it can flex and the windows open slightly so that hydrogen can enter. As the pressure is raised to 1 bar, roughly atmospheric pressure, more hydrogen stuffs into the cavities, attaching itself to the scaffold and making it more rigid. This forces the windows to close, trapping the hydrogen inside. At this point, the pressure can be reduced to as low as 0.01 bar and the inherited rigidity keeps the hydrogen trapped.

This all takes place at the temperature of liquid nitrogen, but it means that when the temperature is raised, the windows can open again and set the hydrogen free. The system is not perfect by any means, admits project scientist Darren Bradshaw, at the University of Liverpool, noting the low storage level achieved so far: just 1% by weight. But, he says, "it's a proof of principle that small windows that dynamically open are a good way to store hydrogen".

If MOF technology is going to be a serious contender for hydrogen storage, industry needs to get involved, not least for scaling-up produc-



Cheaper, cleaner: compressed natural gas is becoming an increasingly popular fuel source worldwide.

tion. MOFs are not yet cheap, easy things to make. They also contain metals, which can be expensive. This is where BASF comes in.

BASF has been involved in MOF research for eight years and has switched its activities from the research division into BASF's daughter company, BASF Future Business in Ludwigshafen, Germany. "The transfer from academia to industry is really a tough job," says

BASF project manager Thomas Danner. Researchers need only small amounts of product, so the solvents and starting materials can be more expensive than those used in industry.

In the short term BASF is not looking at MOFs for storage of hydrogen, but rather for natural gas or methane. "So many questions remain to be answered for the hydrogen economy," says Danner. Natural gas, however, is an existing market, with millions of vehicles

running on compressed natural gas (CNG) already on the roads. Today, most CNG vehicles are used for public transport, and buses can tolerate the bulky storage tanks and shorter driving distances. But as the number of CNG vehicles grows, BASF predicts that longer driving distances and lighter storage will become key demands.

More importantly, MOFs can trap complex gases, such as methane, more easily than they can hydrogen. These gases have more electrons available to interact with both the organic and

metal groups. Existing MOFs can store methane and carbon dioxide at room temperature, circumventing one of the biggest challenges of hydrogen storage.

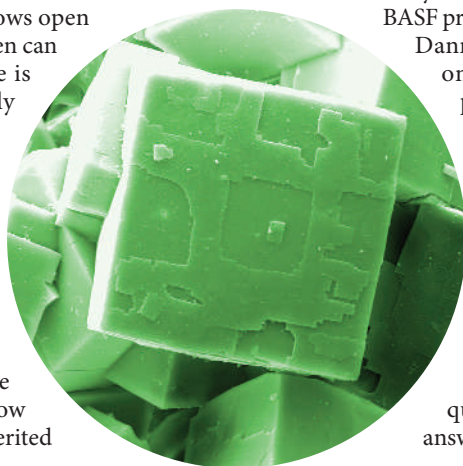
Scaling up

BASF can now make kilograms at a time of MOFs. Danner says that the development stage will last until at least 2009, at which point he expects to see a MOF product on the market. The challenge is not necessarily the technology, but finding the right market for the right product. According to Danner, adding MOFs to a standard CNG fuel tank operating at 200 bar pressure can already increase the distance travelled with a single tank by 25%, but he sees no reason why this cannot be increased to 45% in future.

For now, the practical potential of these super-sponges seems as vast as the space they contain. And whether MOFs ultimately deliver on their gas-storage promise, or disappoint, as carbon nanotubes did, researchers will find other uses for them. For example, their ability to absorb large amounts of guest molecules makes them ideal for catalyst applications. Férey remains amazed that almost no research has been done on MOF catalysis. He is also developing a biomedical MOF for delivering vast quantities of drugs to the bloodstream, while Yaghi is pursuing carbon dioxide storage in MOFs. "It's the very beginning, the limits are in our imagination," says Férey.

Katharine Sanderson is a reporter for Nature based in London.

1. Yaghi, O. M., Li, H., Davis, C., Richardson, D. & Groy, T. L. *Acc. Chem. Res.* **31**, 474–484 (1998).
2. Férey, G. *Science* **309**, 2040–2042 (2005).
3. Furukawa, H., Miller, M. A. & Yaghi, O. M. *J. Mater. Chem.* **17**, 3197–3204 (2007).
4. Lin, X. *et al. Angew. Chem. Int. Edn* **45**, 7358–7364 (2006).
5. Zhao, X. *et al. Science* **306**, 1012–1015 (2004).



How solid is this? The cavities inside metal-organic frameworks can easily adsorb gases.

A. QURESHI/AFP/GETTY BASF

Scientific bodies must take own action on emissions

SIR — Many of the world's most reputable and best-placed scientific organizations, including the American Association for the Advancement of Science, the Royal Society, the American Geophysical Union and the American Meteorological Society, have released strong and unequivocal statements regarding the dangers the world's population faces as anthropogenic climate change gains pace. Although such statements are effective in informing public opinion and thereby influencing policy on this important issue, they are not the most powerful means available.

A more potent approach would be for scientific organizations to make ambitious, high-profile moves to reduce their own contributions to climate change. Such activity could generate significant publicity and demonstrate that the organizations are taking the threat of climate change seriously. They would send a louder, clearer message that emissions reduction should be a priority.

Such moves, although necessarily bold, should not impair the organizations' abilities to achieve their primary aims. Rather, they should publicly demonstrate that reductions in any organization's environmental impact need not reduce its effectiveness. One example would be the more widespread inclusion of video-conferencing facilities in oral sessions at scientific meetings. Another could be the introduction of 'virtual poster sessions' with live audio connections.

If well-implemented, such measures would actually increase the effectiveness of a meeting while reducing its environmental impact. In particular, those who would otherwise not attend could now participate, which would lead to an increase both in the dissemination of research findings and in the interaction between members of the organization.

Andrew Biggin

Paleomagnetic Laboratory Fort Hoofddijk,
Faculteit Geowetenschappen,
Universiteit Utrecht, Budapestlaan 17,
3584 CD Utrecht, the Netherlands

Bush has not obstructed environmental protection

SIR — Your recent Editorial 'An end in sight' (*Nature* **447**, 886; 2007), is completely off base in saying that the US Environmental Protection Agency's mission has been obstructed by the Bush administration. In fact, the EPA's enforcement programme is stronger than ever. The agency was also voted one of the ten best places to work in the federal government, according to

the Partnership for Public Service, an organization based in Washington DC that aims to establish the federal government as an attractive employer for top workers.

During the past three years, EPA's enforcement programme has prevented nearly 1.3 million tonnes of pollution from entering the environment. Its enforcement actions have required defendants to invest \$19.7 billion to reduce pollution at facilities, clean up spills, install new pollution-control equipment and clean up contaminated soil or water. And thanks to the efforts of a revitalized criminal enforcement programme in 2006, offenders will serve a total of 154 years in jail and pay almost \$43 million in fines, as well as another \$29 million for environmental projects imposed as part of their sentences.

Granta Nakayama

Office of Enforcement and Compliance
Assurance, US Environmental Protection Agency,
Washington, DC 20460, USA

Chemical reaction to the many-worlds hypothesis

SIR — Your Editorial 'Parallel worlds galore', celebrating the fiftieth anniversary of the multiverse hypothesis (*Nature* **448**, 1; 2007), was a timely and stimulating reminder that the interaction of science and fiction sometimes leads to facts, or at least to scientific implications. But in your two News Features on the topic (*Nature* **448**, 15–17 and 18–21; 2007), I wonder why chemistry was largely ignored, as this discipline offers excellent scientific platforms for science-fiction narrative. Likewise, fiction holds lessons for chemistry.

Chemists have always wondered whether life in a parallel universe could actually be based on atoms other than carbon. Perhaps that thinking inspired Lewis Carroll's Alice to call into question whether milk could be different in the universe behind the mirror in *Through the Looking Glass*.

Although Kurt Vonnegut's *Galapagos* is quoted as a great example of science fiction in biology, his 1963 novel *Cat's Cradle* has lessons for science. Here, the protagonist, Felix Hoenikker, creates a solid form of water called ice-nine — a narrative probably influenced by the work of Nobel laureate Irving Langmuir. At least 12 forms of ice have now been reported, including ice IX — though this is unlike the fictional ice-nine. *Cat's Cradle* also inspired interest in what seemed to be an extraordinary form of water with anomalous properties, known as 'polywater'. This Russian discovery may have had its roots in papers published during the late 1920s on the vapour pressure of water and other liquids in small capillaries (see, for example, J. L. Shereshfetsky *J. Am. Chem.*

Soc. **50**, 2966–2980 and 2980–2985; 1928). But the research caused much excitement when it was presented in the West in 1966, and it became widely accepted until proved, a few years later, to be fallacious.

It seems appropriate, therefore, to consider chemistry as a tool for uncovering facts that either inspire or imitate fiction. What is really astonishing, to a chemist, about a parallel world coexisting with our own is not its existence in itself, but rather the matter we could find there.

Pedro Cintas

Department of Organic and Inorganic
Chemistry, University of Extremadura,
E-06071 Badajoz, Spain

International research may leave women adrift

SIR — The difficulties facing women and under-represented minorities in science and engineering are highlighted in your Naturejobs Special Reports 'Beyond the glass ceiling' and 'Closing the gender gap' (*Nature* **448**, 98–100 and 101–102; 2007). But women from developing countries who work abroad face additional problems, which are not addressed by efforts to help either women or minorities.

During our studies and beyond, we face financial troubles, assumptions of ineptitude by faculty members, harassment and lack of mentoring. In a survey I conducted among 13 women who graduated from the Indian Institutes of Technology and went to the United States, several reported problems with supervisors, including harassment or having to switch supervisors; two of them eventually left research because of lack of support. On top of that, the stress caused by problems with elderly family members at home can be enormous and is not always understood by people in Western societies, whose family ties and responsibilities may differ. Visa problems are also not unusual.

For those of us who return to our home countries, there are more hurdles to clear. Institutions in many developing countries do not open their doors easily to women, even if they actively seek out and recruit men who have studied abroad. Often it is hard to find a position that is appropriate for our education and experience. And the styles of working abroad and at home can be very different.

To make the best use of our talents and training — whether in the United States, Europe or elsewhere — a full survey of our concerns should be conducted and used to frame new policies and instil change within academic culture.

Asha Gopinathan

GenSci-e-Tech, E 2 Sree Vilas Lane, Kaudiar,
Trivandrum 695003, Kerala, India
dendron.15@gmail.com

BOOKS & ARTS

The art of persuasion

Surprisingly, the rhetoric of the literary artist still has a place in persuasive scientific texts.

The Scientific Literature: A Guided Tour

edited by Joseph E. Harmon and Alan G. Gross

University of Chicago Press: 2007. 312 pp.

\$29.00 (pbk); \$72.50 (hbk)

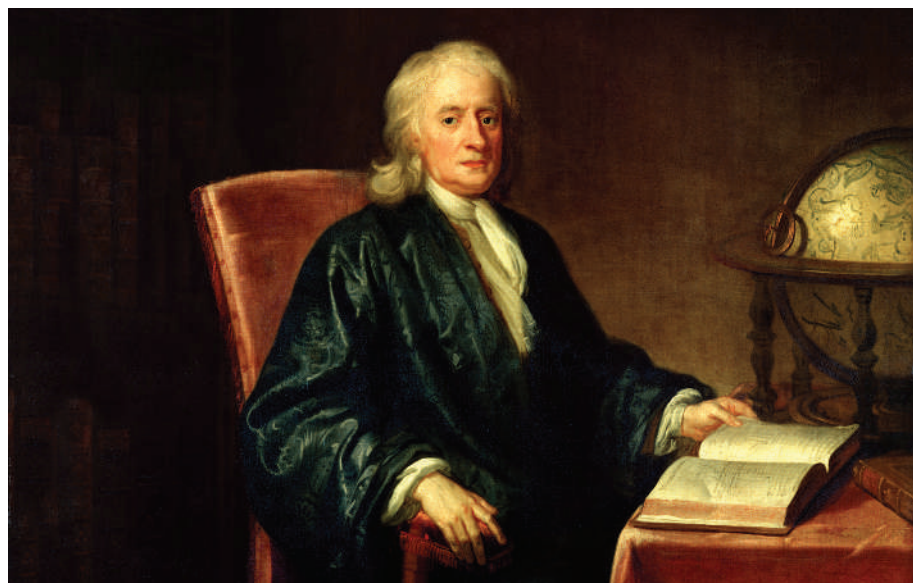
Steven Shapin

While the term 'scientific literature' is a commonplace usage, few scientists would acknowledge any connection between how they write and the works of novelists or poets. As long ago as the middle of the seventeenth century, the English originators of the scientific journal vigorously set themselves against all forms of fancy writing. The newly formed Royal Society of London separated "the knowledge of Nature...from the colours of Rhetorick". The aim of scientific writing was to report, whereas rhetoric worked to distort. Today, few scientists consider themselves to be rhetoricians. How many even know the meaning of anaphora, antimetabole or litotes?

But it's not that simple. The scientific literature reports, but it also aims to persuade readers that what it reports is reliable and significant. And the arts of persuasion are inevitably literary and, specifically, rhetorical. It is an arduously learned skill to write in the way that *Nature* deems acceptable. Conventions of scientific writing have changed enormously over the past few centuries and even over recent decades. The very big differences between Jane Austen's *Persuasion* and a scientific paper lie in the different patterns of rhetoric used in the latter, not in their absence from it.

There are now many historical and sociological studies of scientific communication. Joseph Harmon and Alan Gross's book, *The Scientific Literature*, is something different — neither a research monograph on the history of scientific writing nor a straightforward compilation of excerpts. Originating from an exhibition held at the University of Chicago in 2000, it includes about 125 examples of scientific writing taken from papers, books, reviews and Nobel speeches, and covers material from the seventeenth century up to the announcement of the rough draft of the human genome in 2001.

An excerpt is rarely longer than 500 words and sometimes as brief as 150, or may just be a diagram. These scientific snippets are embedded in strands of editorial commentary describing, highlighting and interpreting them. The tone is genial: this "guided tour" doesn't threaten arduous intellectual



Robert Boyle (1627–1691) sometimes used the first-person singular in his scientific writing.

adventure. Rhetorical terms are explained, scientific authors are identified, and pertinent scientific contexts introduced.

There is no single argument embodied in this book — more a selection of sensibilities intended to help readers appreciate the remarkable and shifting set of literary forms that scientific writing has assumed. One theme is historical change. The authors point out that, not surprisingly, specialization has been accompanied by increasingly exclusive scientific writing. There never was a golden age when every educated person could read everything in the scientific literature — Newton's *Principia* defeated all but a small number of natural philosophers and mathematicians. But until the mid-nineteenth century, the general readership of such periodicals as the *Edinburgh Review* might find serious treatments of what was up in geology, astronomy or mathematics, written by notable scientists.

The accelerating incomprehensibility of scientific writing to the average educated person is not merely the fault of the much-lamented 'public ignorance of science'. Specialists have been so successful in constructing and bounding their own audiences that they rarely feel any need to address the laity or even scientists in other disciplines. Indeed, the plant physiologist is likely to be just as poorly equipped as any non-scientist to read a paper on superconductivity.

Another theme is the impersonality of scientific prose. Scientific writing has always been relatively impersonal, but the literary forms of impersonality have changed over time. In the seventeenth century, Robert Boyle used thickly layered circumstantial reporting to portray himself as a modest witness of his experiments, his judgement uncoloured by theoretical interest. He was nevertheless a witness at the centre of his own narratives, not averse to using the first-person singular — "I did X, I saw Y". By the nineteenth century — when the French physiologist Claude Bernard coined the aphorism "Art is I; Science is We" — the scientific author became increasingly submerged in either the first-person plural ("We did X, we saw Y") or in the passive voice now standard in scientific papers ("X was done, Y was seen").

The rhetorical convention here implies that scientific authors do not matter to what they report in the same way that Jane Austen matters to *Persuasion*. Although some insist that scientific research is an imaginative exercise and that its findings have an aesthetic character, the convention of impersonality is testimony to the opposite sensibility. Science is considered to discover; art to create.

Harmon and Gross are quite right to draw attention to non-verbal forms of communication and the changes produced by both instrumental and representational technologies on the ability of the scientific literature to show as

well as say. Wood or copperplate engravings were important in seventeenth-century science, but such images were expensive to produce and limited in their information content. Now, practically every issue of a scientific journal is a cornucopia of high-bandwidth visual communication sometimes even in online video form. It is becoming easier to envisage present-day science communication without words than

without images. It is disappointing then that many of the illustrations in *The Scientific Literature* are so murkily reproduced. Maybe it is easier for humanists to say that visual communication is important than for them and their publishers to act as if it is. ■

Steven Shapin is in the Department of the History of Science, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, USA.

Material metaphors

Origins and Revolutions: Human Identity in Earliest Prehistory by Clive Gamble

Cambridge University Press: 2007. 362 pp.
\$80, £45 (hbk); \$27.99, £15 (pbk)

Robert N. Proctor

Research into human origins can be thought of as a kind of identity quest. We want to know how 'they' became 'us', which raises all kinds of questions about what it means to be human. To stand upright? To paint the walls of caves or to fashion beads from bone? Or to plant the land and build cities with slave labour? Or perhaps to engage in none of the above, but simply to have that capacity?

Questions such as these do not have obvious answers, nor are they really even empirical. Evolution stretches out the process of anthropogenesis. Once we jettison teleology and discontinuity, it doesn't mean much to say when hominins became 'truly human', any more than to say when aardvarks became truly aardvark. Nor can it even mean much to talk about the 'earliest' humans, or prehistory, as everything will depend on what we want to identify as the important transitions.

Upright posture, for example, appeared by about 4 million years ago, but tool making must be much older, albeit invisible as a result of accidents of preservation. The oldest known wooden tools, the famous spears from Schöningen in Germany, date from only 400,000 years ago. Symbolic burial and bead making are younger still, perhaps by an order of magnitude.

Clive Gamble's new book, *Origins and Revolutions*, challenges our current obsession with language and farming as the two principal 'big breaks' in deep antiquity, dating from around 40,000 and 10,000 years ago, respectively. His intention is to avoid all talk of origins, exploring instead what he calls the "material basis of human identity", by which he means how artefacts as extensions of the human body acquire a symbolic force of their own.

He divides the material world into "instru-

ments" and "containers". Here instruments include all edges, blades and points, as well as pestles, ploughs, drills, axes, brushes, writing implements and wheels. Containers include anything that houses or envelops, whether in the form of bowls, barns, bags, caves, clothes, moulds, masks or tombs. Gamble's point is that both are extensions of the human body: instruments extend our limbs; containers extend our trunk. Instruments generally inscribe; containers are more often inscribed upon.

Classifying material culture in this way



Tools (harpoons, awl and needle with eye) from the Upper Palaeolithic.

allows Gamble to question the novelty of both the Neolithic and the Upper Palaeolithic transitions. He argues instead for a more gradual shift over millions of years of hominin evolution, from a life centred around instruments to a life more prominently incorporating containers. Farming, then, is not such a radical innovation. There is no 'sapient paradox' — Colin Renfrew's puzzle over why it took so long to discover agriculture and the virtues of a sedentary life. Symbolism was not suddenly invented when modern humans decided to quit Africa and start painting in southern France. (Paradox seekers might well wonder why the 'modern mind' seems to appear 100,000 years after the 'modern body'.) Instruments and containers always reference the human body, and in this sense carry symbolic force. This means that symbolism does not necessarily have a

singular point of origin, whether 40,000 years ago with the 'human revolution' or at any other magical moment.

Palaeoanthropology has become an exciting field in recent years, partly because some really big questions remain wonderfully unanswered. No one really knows whether Neanderthals could speak or think like us, for example, or what it might have been like to live among our *Homo erectus* next-of-kin or the newly discovered *Homo floresiensis*. In the 1960s and 1970s, language, art and symbolism were projected onto ever-older hominin fossils. Now the trend is to (re-) dehumanize early palaeolithic hominins — hence the darkening of the whites of their eyes in recent museum displays.

Gamble's refocus on instruments and containers is a refreshing break from archaeological convention. But how far back can we go before such proxies for the hominin body cease to have symbolic force? Birds' nests are containers, so when do the hominin counterparts start to signify something more to their makers? When do the instruments of early hominins start to serve as material metaphors? How would we ever know whether, say, the invention of symbolism wasn't rather sudden,

even from a geological point of view? How would we ever know whether a light went on in some hominin head (or gene), causing language to spring into being?

Gamble shows that the rate of invention grows slowly over the long haul of human evolution, and reminds us that absence of evidence is not evidence of absence. But how long should we search the Middle Palaeolithic for painted caves or sculpted figurines before concluding that none was ever done, and not for lack of interest, but for lack of capacity? For many years, geologists were reluctant to recognize catastrophes, postulating 'missing strata' to account for apparent jumps. The

rehabilitation of catastrophes over the past few decades owes much to a renewed appreciation that absence of evidence can be evidence of an absence. I think it is fair to ask whether the situation might not be similar for palaeoanthropology.

Origins and Revolutions is an effervescent read that skillfully challenges many of the sacred cows of archaeology. It is rich and deep in the philosophical acumen and attention to social theory for which Gamble is known. He also writes with an admirable sense of humour and irony; he knows how to join humanistic flair with empirical rigour at the dig.

I think he is right that our bodies are a kind of social technology, and that artefacts should be regarded as embodied metaphors. The question then arises of how to understand changes in interactions between artefacts, with

the invention of compound tools. Following Gamble, compound artefacts might well be thought of as material metaphors for language. They help make metaphors, and in this sense language, possible.

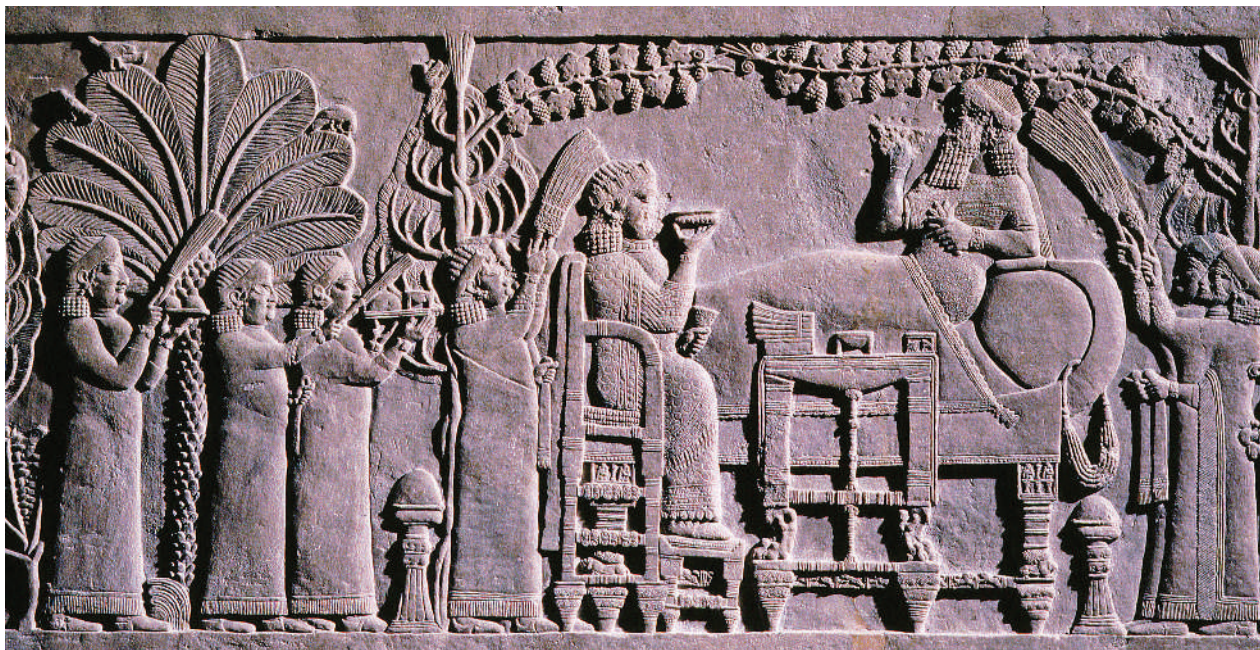
Compound artefacts open up a near-infinite

recombination of uses, just as language allows a near-infinite recombination of words. If there's any truth to the idea that language and composite tools arose together, surely we can ask how big or how sudden must such a change have been to constitute a revolution. As Gam-

ble himself asks, when challenging stone-flaking taxonomies: "At what moment does eating produce the core of an apple?" ■

Robert N. Proctor is professor of the history of science at Stanford University, Stanford, California 94305, USA.

A. DAGLI ORTI/BRITISH MUSEUM/ THE ART ARCHIVE ARCHIVE



Left to digest

In ancient art, banqueters always recline on their left side — perhaps to aid digestion.

Paolo Mazzarello and Maurizio Harari

The élite of most advanced ancient Mediterranean societies partook of banquets lying down. We know this from iconographic records dating back to the seventh century BC. Some scholars assume that the custom was widespread in the originally nomadic tribes that finally settled in Syria or Iran, befitting their modest tent furnishings. But the social prestige that soon became associated with reclining at a banquet might owe more to the preciousness of the beds of the rich, as suggested by the outpouring of the biblical prophet Amos (around 750 BC), against those used by the Samaritans: "Lying upon beds of ivory, stretched comfortably on their couches, they eat lambs taken from the flock." (*Bible*, Amos 6:4–7).

So it's not surprising that one of the oldest images of a reclining banquet is a royal one: the famous bas-relief of King Assurbanipal of Assyria lying on his left side while his wife sits on the throne (pictured). This form of aristocratic banquet was widespread in the seventh century BC in Greece — the poet Archilochus wrote, "leaning on my lance I drink (wine)" — and among the Etruscans,

who traded with the Greeks. It came to span the entire Mediterranean Greek and Roman civilizations.

Art historians have often noted that banqueters almost always appear to be reclining on their left sides. The usual explanation is that lying on the left leaves the right hand free to hold the dining vessels. But in funereal art there is good documentation of presumptive left-handed banqueters also reclining to the left. Jean-Marie Dentzer in his book *Le motif du banquet couché dans le Proche-Orient et le monde Grec du VIIIe au IVe siècle avant J.-C.* (Ecole Française, Rome, 1982) has compiled an extensive inventory of the *banquet couché* between the seventh and fourth centuries BC. Of the more than 700 illustrations, including at least a dozen banqueters holding pots in their left hand, not one is lying on their right side.

One explanation could lie in the anatomy of the stomach and in the digestive mechanism. The stomach has an irregular shape that curves upon itself. Its rounded base is turned to the left. There are two openings: one at the top where food enters from the oesophagus and one at the base,

the pyloric orifice, from which part-digested food exits.

Eating lying down increases abdominal pressure and thus promotes gastro-oesophageal reflux, in which the acid stomach contents are forced back up the oesophagus, causing the unpleasant sensation known as heartburn. When lying on the left, the chewed food has room to expand because the curvature of the stomach is enhanced in that position. The lesser curvature on the right side of the stomach gives little space for food to resist the increased abdominal pressure and so lying on the right will soon cause reflux.

For the ancient Mediterranean civilizations, the evening meal lasted hours, and involved a lot of eating and drinking. Lying on the left would not just have reduced the risk of reflux, it would also have provided space for the large amounts of food the revellers were required to eat. ■

Paolo Mazzarello is professor of humanities and the history of medicine in the Department of Experimental Medicine, Maurizio Harari is professor of Etruscology in the Department of Classics, University of Pavia, 27100 Pavia, Italy.

SCIENCE IN CULTURE

BIOLOGICAL CHEMISTRY

Enzymes line up for assembly

Nicholas M. Llewellyn and Jonathan B. Spencer

Many enzymes have a series of catalytic sites, lined up like beads on a string. A previously unknown link in one of these molecular assembly lines involves an unexpected approach to a common biochemical reaction.

Nearly 100 years ago, Henry Ford demonstrated the full strength of economist Adam Smith's insights into productivity and the division of labour when he established the first moving assembly line. By shuttling partially constructed cars mechanically from one worker to the next, each performing a single specific task, Ford's assembly line could issue a new Model T every three minutes. This manufacturing method provided the foundation of modern mass production. But nature employed much the same approach for constructing molecules long before humans existed to ponder questions of economy and efficiency. On page 824 of this issue, Walsh and colleagues¹ identify a previously unrecognized link in one such biological assembly line — an enzyme that could some day be exploited by chemists to modify complex, naturally occurring compounds.

The enzymes that form the polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) families are responsible for the biosynthesis of many useful compounds, including the antibiotics erythromycin and vancomycin, and the antitumour drug epothilone. These multi-subunit enzymes are the molecular equivalents of moving assembly lines: growing substrate molecules are handed, bucket-brigade style, from one specialized catalytic site to the next, with each site performing a specific and predictable function (Fig. 1)^{2,3}. The catalytic domains that make up these complex biosynthetic machines are so well studied that the likely product of a newly discovered PKS or NRPS gene cluster can often be predicted from the gene sequence alone.

The PKS assembly line starts by recruiting small building-blocks (such as acetate and propionate molecules, which contain 'acyl' chemical groups) onto carrier proteins. The building-blocks are then bonded together in reactions catalysed by a 'ketosynthase' region of the PKS. The resulting substrate may then be chemically tailored by various other enzyme domains, before being passed on to another ketosynthase for a further round of extension and modification. The cycle is repeated until the finished molecule is finally offloaded. The various catalytic domains may exist as discrete enzymes (as in type II PKS), or be

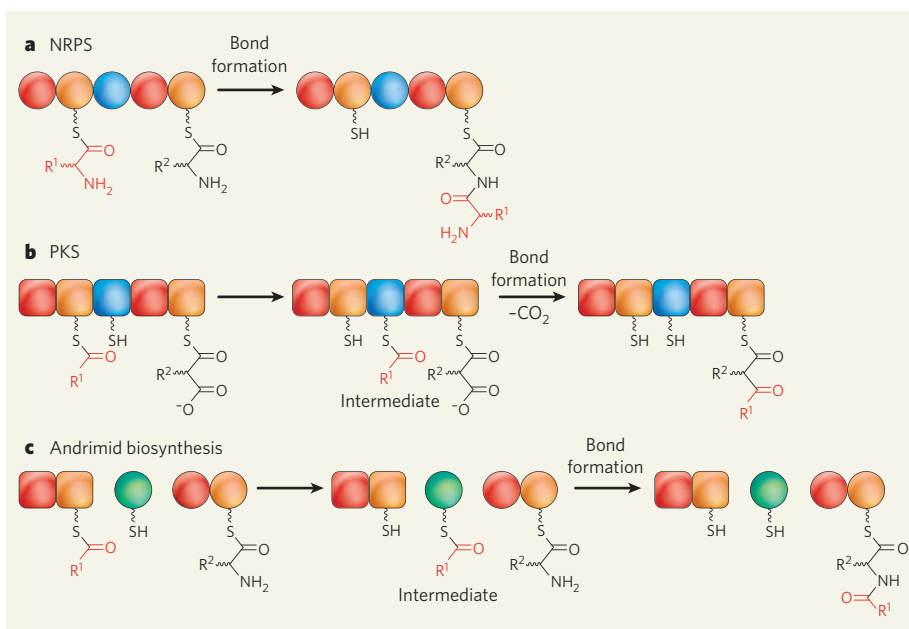


Figure 1 | Assembly-line logic in biosynthetic enzymes. **a**, Non-ribosomal peptide synthetase (NRPS) enzymes construct peptides from amino-acid substrates using an assembly line of domains. Each domain performs a specific task — substrate recruitment (red), substrate carriage (orange) or bond formation (blue). The bond-forming domain aligns the substrates ready for reaction, but does not covalently bind to the substrates. **b**, Polyketide compounds are prepared by polyketide synthase (PKS) enzymes, using the same strategy as above, but with different substrates and protein domains. Bonds between substrates are formed via an intermediate in which the substrate is covalently attached to a bond-forming domain. **c**, Walsh and colleagues¹ show that in the biosynthesis of the antibiotic andrimid (a hybrid peptide–polyketide compound), the bond-formation reaction is performed by a previously undiscovered enzyme (green). The resulting bond resembles those found in NRPS products, but its mechanism of formation involves a covalently bound intermediate, as for PKS enzymes. R¹ and R² represent any hydrocarbon group.

connected end to end, like beads on a string (as in type I PKS), but in both cases the biosynthetic strategy remains the same.

The NRPS cycle is very similar to that of PKS enzymes, except that it uses amino acids as building-blocks. Thus, amino acids become bound to peptidyl carrier proteins (PCPs); PCP-bound amino acids are joined together with amide bonds to form peptides, in catalytic sites known as condensation domains; tailoring regions may then modify the newly formed peptide before passing it along for further cycles of extension and tailoring; and finally, the finished product is cleaved from the enzyme. The PKS and NRPS enzymes each produce very

different products, but the logic they use is strikingly similar — so similar, in fact, that they can easily cooperate to construct hybrid PKS–NRPS products such as epothilone⁴.

Andrimid, an antibiotic that targets fatty-acid biosynthesis in bacteria, is another of these hybrid products. It was first isolated and its structure elucidated⁵ in 1987; the gene cluster associated with the biosynthetic pathway for andrimid was then identified and sequenced⁶ in 2006 by Clardy and co-workers. Although much of the PKS and NRPS machinery responsible for assembling andrimid was easily identified in this cluster using computational techniques, the essential NRPS condensation domains could

not be found. How, then, were the two amide bonds in the andrimid skeleton formed?

Clardy proposed that the proteins AdmF and AdmS might fulfil the roles of the missing condensation domains. These enzymes are similar to transglutaminase proteins, which crosslink peptides in a process that breaks amide bonds in the side chains of glutamine amino acids, and then re-forms them with other amino acids. After a thorough *in vivo* and *in vitro* investigation, Walsh and co-workers¹ now confirm that AdmF is responsible for the formation of the first amide bond in andrimid biosynthesis.

The mechanism of AdmF catalysis proceeds through a covalent acyl-enzyme intermediate (Fig. 1c). This is more similar to the mechanism used by PKS ketosynthase domains than the direct amide coupling performed by NRPS condensation sites. The authors' discovery¹ thus reveals an alternative amide-formation process to that of established NRPS condensation domains, and uncovers a further option in the NRPS tool-kit of reactions. In addition, Walsh and colleagues identified genes that encode analogous enzymes to AdmF and AdmS in gene clusters in other bacteria, suggesting that this unusual amide-formation strategy is not unique to andrimid biosynthesis.

Henry Ford is famously reputed to have said: "Any customer can have a car painted any colour that he wants so long as it is black." Similarly, the finely tuned PKS and NRPS assembly lines tend to make only one product, with modest variation at best, despite their tantalizing potential to produce libraries of analogous compounds from diverse pools of building-blocks. AdmF, however, is reported to tolerate several kinds of acyl substrate *in vitro*, suggesting that manipulation of its biosynthetic feedstock could allow the production of a variety of andrimid analogues. A few other catalytic sites from PKS and NRPS enzymes — such as PikAV, which forms part of the biosynthetic machinery for the antibiotic pikromycin⁷, and the discrete CytC1 domain from γ -dichloroaminobutyrate synthetase⁸ — are also reported to accept several substrates. These relatively promiscuous domains might one day be exploited by chemical biologists attempting to prepare analogues of naturally occurring potential drugs, in the hope of finding new medicines.

Nicholas M. Llewellyn and Jonathan B. Spencer are in the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. e-mail: jbs20@cam.ac.uk

- Fortin, P. D., Walsh, C. T. & Magarvey, N. A. *Nature* **448**, 824–827 (2007).
- Staunton, J. & Weissman, K. J. *Nat. Prod. Rep.* **18**, 380–416 (2001).
- Finking, R. & Marahiel, M. A. *Annu. Rev. Microbiol.* **58**, 453–488 (2004).
- Cane, D. E. & Walsh, C. T. *Chem. Biol.* **6**, R319–R325 (1999).
- Fredenhagen, A. et al. *J. Am. Chem. Soc.* **109**, 4409–4411 (1987).
- Jin, M., Fischbach, M. A. & Clardy, J. *J. Am. Chem. Soc.* **128**, 10660–10661 (2006).
- Akey, D. L. et al. *Nature Chem. Biol.* **2**, 537–542 (2006).
- Ueki, M. et al. *Chem. Biol.* **13**, 1183–1191 (2006).

SEISMOLOGY

Talc at fault

Christopher Wibberley

The behaviour of the San Andreas fault varies along its length — it slips in some places and creeps in others. The discovery of the ultrasoft mineral talc in rocks from deep inside the fault could help to explain why.

Faults are puzzling beasts. The strength of faults, the largest of which are boundaries between Earth's tectonic plates, has been mired in controversy ever since plate tectonics properly arrived on the scene in the 1960s. And no plate boundary is more puzzling than that of the world's most famous fault, California's San Andreas. Simple experiments on the mechanics of sliding rocks show that the San Andreas should be strong, resisting movement. But the lack of any clear sign along the fault of heating generated by frictional sliding, coupled with the orientation of the compressive stress driving that sliding, are indicative of a much weaker fault^{1,2}. Furthermore, different parts of the San Andreas behave differently: whereas its northern and southern sections 'stick and slip', causing infrequent large earthquakes with little movement in-between times, the central section 'creeps' — it moves little, but often.

On page 795 of this issue, Moore and Rymer³ present surprising findings from a literally ground-breaking project that might help to explain this aberrant behaviour. The project is SAFOD, the 'San Andreas Fault Observatory at Depth'⁴, which has been drilling into the central, creeping section of the San Andreas fault zone since 2003. It aims to determine the nature of the deformed rocks and the

physical and chemical conditions in and around the fault zone 3 km below the surface, and to monitor, *in situ* and over the long term, seismicity, fluid pressure, stress and fault slip (Fig. 1). What Moore and Rymer find in fractured and sheared rocks sampled from the central San Andreas that is so intriguing is talc — the softest known natural mineral.

Talc, a hydrous magnesium silicate, is not a significant constituent of Earth's crust. But hydrothermal fluids circulating in the San Andreas fault have encouraged metamorphic reactions that form talc from magnesium-rich serpentinite rocks, relic fragments of oceanic lithosphere (crust and upper mantle) exhumed by tectonic processes. The presence of talc brings the range of possible frictional fault strengths to a new low — low enough to explain the apparent weakness of the San Andreas today, without the need to invoke long-term, curiously high fluid pressures.

Moreover, in laboratory tests, talc is more resistant to movement when sliding velocity is increased. Scaled up to real-Earth conditions, talc's properties are likely to encourage slow, stable creep, inhibiting the build-up of elastic energy in the volume around the fault and preventing the faster sliding and unstable slip of a large earthquake⁵. So it is probably no coincidence that the creeping part of the San Andreas fault in central California is where these serpentinite rocks are found.

But this discovery does not explain all aspects of the behaviour of the San Andreas' central region. Why, if talc slides stably, do earthquakes occur there at all? The initiation of an earthquake requires the nucleation of unstable motion, including rock resistance that weakens with increasing velocity — the very opposite of talc's behaviour. The most likely explanation is that the fault zone is very complex, cutting through different rock types and containing various kinds of altered material, as well as several slip zones surrounded by deformed rock in a region several metres or more wide^{6,7}. A rupture front that starts in a non-serpentinite patch of the central San Andreas may thus quickly die out when it passes through serpentinite if talc is present there. Furthermore, some of

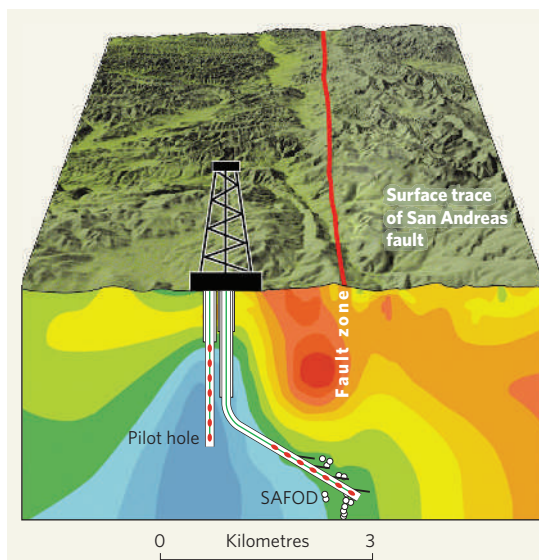


Figure 1 | In search of weak creep. A schematic cross-section of the San Andreas fault at Parkfield, California, site of the SAFOD drill hole for investigating areas of persistent minor seismicity (white dots). The subsurface colouring indicates the electric resistivity of the rocks; particularly low-resistivity areas (red) could indicate fluid-rich zones of the type needed for the formation of talc, as found by Moore and Rymer³.

the small earthquakes seen in the region may arise at geometric irregularities or at secondary fractures away from the main slip zones, again a result of the complex internal structure of the fault zone⁸.

A second unresolved problem is that, in the classic view, the strength of Earth's crust increases with depth in the upper to middle crust owing to the weight of the rock above, until a point between about 10 and 15 km down, where temperature comes into play. The weakening issue should therefore be most serious for the middle crust, where the predicted strength is greatest. Along the creeping part of the San Andreas, where serpentinite is known to be present at the relatively shallow depths of 3–7 km, deeper weakening processes must also be operating to fully explain the deviation of the fault from classic strength profiles. Our current knowledge of the distribution of serpentinite rocks does not, however, allow us to propose with confidence that talc is present at these deeper levels. It also indicates that other factors complement the weakening effect of talc at shallow depths.

What is more, along the central San Andreas the lack of seismicity lower down towards the middle crust — where large earthquakes on other parts of the fault often nucleate⁹ — indicates that stable slip is occurring, probably through viscous deformation. This depth range is usually dominated by unstable, frictional mechanisms. In the classic view of crustal strength, viscous deformation mechanisms (such as crystal plasticity, and creep owing to the dissolution and subsequent recrystallization of rock grains under high-pressure conditions) do not operate at depths shallower than 10–15 km because temperatures above this level are insufficient.

But geological studies of exhumed former strands of the San Andreas fault system, and other similar ancient faults such as the Median Tectonic Line in Japan, show that strongly aligned phyllosilicate minerals such as micas and clays (and including talc), which are viscous at lower temperatures, replaced stronger minerals (feldspar, for example) in the fault zone at these considerable depths. This process was caused by localized metamorphic reactions encouraged by fluids within the fault zone, and is thought to explain the weak, aseismic behaviour of large faults at greater depths in the middle of the continental crust even if, as in most cases, serpentinite rocks are not present^{10–12}.

The overall lesson is that fluid-triggered metamorphic reactions localized in fault zones can strongly influence the strength of faults along plate boundaries and their likelihood of generating large earthquakes. Basic geological constraints such as the three-dimensional distribution of rock types and structural complexity in fault zones at relatively small scales are pivotal in the control of fault mechanics and tectonic behaviour at much larger scales. Moore and Rymer's discovery³ is a prime example, and might not be the only gem that the

SAFOD project digs up. Future collaboration between structural geologists and seismologists should yield better definitions of the composition and structure of fault zones, and their influence on the nucleation and propagation of earthquakes.

Christopher Wibberley is at the Laboratoire de Géosciences Azur, Université de Nice Sophia Antipolis, France. (In October 2007 he will move to TOTAL Centre Scientifique et Technique Jean Feger, Avenue Larribau, F-64018 Pau Cedex, France.)

e-mail: wibbs@geoazur.unice.fr
(christopher.wibberley@total.com)

1. Lachenbruch, A. H. & Sass, J. H. *J. Geophys. Res.* **85**, 6185–6222 (1980).
2. Zoback, M. D. *et al. Science* **238**, 1105–1111 (1987).
3. Moore, D. E. & Rymer, M. J. *Nature* **448**, 795–797 (2007).
4. Hickman, S., Zoback, M. & Ellsworth, W. *Geophys. Res. Lett.* **31**, L12501 (2004).
5. Scholz, C. H. *Nature* **391**, 37–42 (1998).
6. Chester, F. M. & Chester, J. S. *Tectonophysics* **295**, 199–221 (1998).
7. Wibberley, C. A. J. & Shimamoto, T. *J. Struct. Geol.* **25**, 59–78 (2003).
8. Unsworth, M. J., Malin, P. E., Egbert, G. D. & Booker, J. R. *Geology* **25**, 359–362 (1997).
9. Sibson, R. H. *J. Geol. Soc.* **140**, 741–767 (1983).
10. Janecke, S. U. & Evans, J. P. *Geology* **16**, 1064–1067 (1988).
11. Wintsch, R. P., Christoffersen, R. & Kronenberg, A. K. *J. Geophys. Res.* **100**, 13021–13032 (1995).
12. Jefferies, S. P. *et al. J. Struct. Geol.* **28**, 220–235 (2006).

BIOCHEMISTRY

Designer enzymes

Michael P. Robertson and William G. Scott

Evolution has crafted thousands of enzymes that are efficient catalysts for a plethora of reactions. Human attempts at enzyme design trail far behind, but may benefit from exploiting evolutionary tactics.

Chemical reactions in living organisms are catalysed by enzymes, the vast majority of which are proteins. These finely tuned catalysts are the result of billions of years of evolution, and far surpass anything yet created by humans. Indeed, our ability to design enzymes, on the basis of our knowledge of protein structure and reaction mechanisms, can most charitably be described as primitive. The structure and catalytic properties of an enzyme are dictated by its amino-acid sequence in ways that are not understood well enough to reproduce. On page 828 of this issue, Seelig and Szostak¹ describe how they bypass this intractable difficulty by simulating evolution. They use an *in vitro* artificial selection process to isolate new protein enzymes that join the ends of two RNA molecules together.

The ability to make enzymes for specific purposes is of great practical interest — designer

enzymes could be made for many potential applications. They could, for example, be used to prepare drugs efficiently. In fact, some methods for preparing new enzymes already exist. One approach is the randomization and *in vivo* selection of variants of existing enzymes. This strategy has been reasonably successful, but it is limited by the relatively small number of possible variants (typically from 10⁶ to 10⁸; for comparison, a system that generates more than 10¹² would be desirable).

Another approach is to use an organism's immune system in a form of natural selection to create catalytic antibodies^{2,3}. Enzymes work by binding and stabilizing the transition state of a reaction — the highest-energy configuration of atoms in the reaction pathway. So if an antibody can bind to molecules that have the same geometry as a reaction's transition state, then it can also catalyse that reaction. Generating catalytic

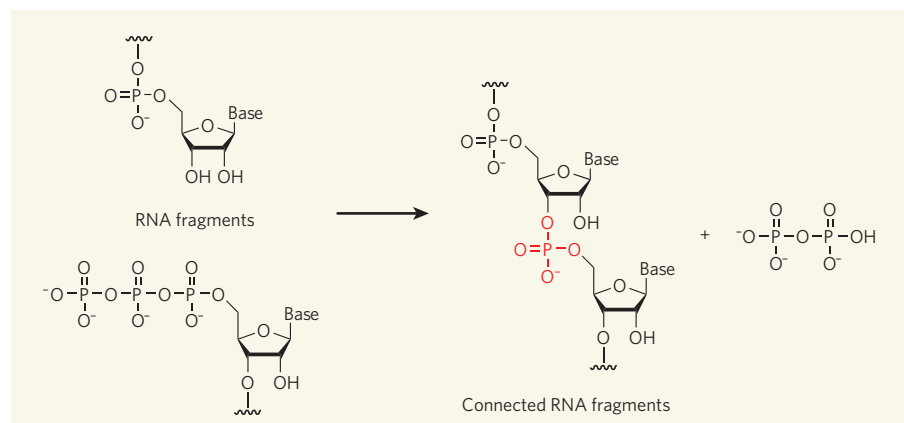


Figure 1 | Artificial enzyme evolution. Seelig and Szostak¹ have developed an *in vitro* method for evolving protein enzymes that can catalyse bond-formation reactions. Using this method, they generated an RNA ligase enzyme that forms a phosphodiester connection (red) between two RNA fragments.

antibodies thus requires a detailed knowledge of the reaction's mechanism and the ability to synthesize a transition-state mimic — conditions that are often not met.

Catalytic antibodies can be thought of as rationally designed enzymes, because knowledge of the reaction pathway is required to make them. But the creation of particular antibodies in this way is purely the product of *in vivo* genetic rearrangements that generate a vast number of antibody variants, and of the immune selection process itself. Catalytic antibodies typically provide a 10^4 -fold to 10^6 -fold rate enhancement of reactions, but usually fall short of the catalytic prowess exhibited by their natural enzyme counterparts. This is probably because transition-state stabilization is only one of several strategies used by natural enzymes to accelerate reactions.

By contrast, enzymes have evolved naturally in a selection process that deals directly with reactants and products, but only indirectly with transition states. Seelig and Szostak¹ use this natural strategy to discover enzymes that catalyse bond-forming reactions. They use a technique known as mRNA display, in which proteins to be screened for catalytic activity are each tethered to the specific messenger RNA (mRNA) that encodes that protein (see Fig. 1a on page 828). Complementary DNA (cDNA) is generated from the mRNA, using a primer that has a substrate attached, thus creating an mRNA–cDNA duplex with a protein and substrate attached. The authors then add a second substrate that incorporates an anchor that can be immobilized. If the protein catalyses a bond-formation reaction between the two substrates, then it will be immobilized — and thus selected — through the anchor. The authors amplify selected cDNA that corresponds to an active enzyme, and use this to encode proteins for a subsequent round of selection. By generating random mutations in the cDNA before amplification, the authors create an artificial evolution process that optimizes enzymes for the catalytic activity over several rounds of selection.

Seelig and Szostak's technique¹ has the practical advantage that the entire process is carried out *in vitro* with a large set of proteins — their study uses 10^{12} unique sequences. The authors used their method to isolate RNA ligase enzymes, which catalyse the formation of a bond between two pieces of RNA (Fig. 1). Previously reported mRNA-display selections⁴ were limited to isolating proteins that simply bind to target molecules. Seelig and Szostak have achieved a significant breakthrough by using mRNA display to evolve proteins that not only bind to target molecules, but also catalyse enzymatic reactions that use the bound target molecules as substrates.

Although proteins have won the fitness contest of natural selection to become the pre-eminent enzymes, billions of years ago life may have started with RNA enzymes — ribozymes — in a putative RNA world that pre-dated proteins and DNA. The RNA bond-forming

(ligation) reaction is a favourite of those studying evolution from an RNA world, because it is presumed to be the crucial chemical step of RNA self-replication⁵. Szostak and fellow molecular biologist David Bartel were the first to isolate a ribozyme ligase⁶, using artificial selection. Their technique is the prototypical method for the *in vitro* evolution of ribozymes, and has been adapted for protein enzymes by Seelig and Szostak in the current study¹.

The authors' protein-based ligases don't make better enzymes than their RNA equivalents, which may seem surprising, given that evolution apparently replaced ribozymes with protein enzymes. So how can this observation be explained? It is possible that some of the catalytic potential of proteins goes untapped by artificial selections that are based only on product formation. mRNA-based selections deal most directly with enzyme–product interactions, whereas earlier catalytic-antibody selections mainly involved interactions between enzymes and transition states. The two methods could potentially

select for different aspects of enzyme catalysis, and so provide complementary glimpses of the processes that governed the evolution of naturally occurring enzymes. Perhaps protein enzymes are better catalysts in part because they have a greater propensity to form transition-state-stabilizing interactions. Designing a selection process that includes ground-state interactions (as Seelig and Szostak's study¹ does) and transition-state interactions (as the previous catalytic-antibody approaches did) might yield even better-designed enzymes. ■ Michael P. Robertson and William G. Scott are at the Center for Molecular Biology of RNA, University of California, Santa Cruz, California 95064, USA.

e-mail: wgscott@chemistry.ucsc.edu

1. Seelig, B. & Szostak, J. W. *Nature* **448**, 828–831 (2007).
2. Jencks, W. P. *Catalysis in Chemistry and Enzymology* (Dover, Mineola, NY, 1987).
3. Schultz, P. G. & Lerner, R. A. *Science* **269**, 1835–1842 (1995).
4. Keefe, A. D. & Szostak, J. W. *Nature* **410**, 715–718 (2001).
5. Joyce, G. F. *Science* **315**, 1507–1508 (2007).
6. Bartel, D. P. & Szostak, J. W. *Science* **261**, 1411–1418 (1993).

MATERIALS SCIENCE

Metal turned to glass

Gilles Tarjus

In order to form a glass by cooling a liquid, the normal process of solid crystallization must be bypassed. Achieving that for a pure metal had seemed impossible — until pressure was applied to liquid germanium.

Glasses are nothing but frozen liquids: closely packed, but randomly ordered assemblages of molecules that no longer flow on any reasonable timescale (by human standards), and thus are solids for all practical purposes. Being a solid while lacking the long-range order typical of that phase of matter is what gives glasses the physical properties that make them so useful in our everyday life. On page 787 of this issue, Bhat *et al.*¹ add another member to this ubiquitous family of materials: a glass made by cooling a pure, 'monatomic' metal.

Glasses can be formed from a great variety of materials, irrespective of the type of interaction prevalent among their constituent atoms or molecules — whether covalent, ionic, hydrogen, van der Waals or metallic bonds. Most commonly, glasses are made by cooling a liquid, with the prerequisite that crystallization into a regular solid does not occur first along the way. More than 50 years ago, David Turnbull discovered^{2,3} that molten metals can be cooled to below their melting point and retain their liquid structure, a phenomenon described as supercooling, or undercooling. But the tendency of such systems to crystallize is very strong, and metallic glasses have so far been obtained only by mixing several cleverly chosen components into an alloy.

Why would finding a one-component metallic glass be so important? It is probably less notable in terms of direct technological applications than for the progress it could generate in our understanding of glass formation. To quote Turnbull⁴: "The most convincing evidence of the universality of the glass state would be the demonstration that pure monatomic substances...can be put into the glass form." One might add that these substances should be characterized by simple inter-atomic forces that do not involve strong directional bonding: liquid selenium, for instance, is a non-metal in which the atoms form covalently bonded, polymer-like chains, and is easily vitrified.

A central issue is to determine the factors that allow crystallization to be bypassed as a substance is cooled past its melting point. It has been known for decades that the rate at which the liquid is cooled must be fast compared with the time taken for crystals to form. The trick is to find ways of slowing the crystallization down. In the case of metallic materials, widely applicable empirical rules have been devised by which the thermodynamic propensity towards crystallization is repressed, and the temperature of crystal formation is reduced. The effect is that atomic motion at the point of

crystallization becomes very sluggish, and the solidifying substance maintains the disorder of the liquid phase — it forms a glass.

Such strategies have been used to find multi-component alloys that can form metallic glasses in bulk^{5,6}. But despite inspiring work by F. C. Franck on sphere packing and 'frustrated icosahedral order'⁷, monatomic liquids have always been found to be extremely poor glass formers. In fact, there was no reported evidence of any one-component metallic glass — until Bhat and colleagues' successful vitrification of metallic liquid germanium¹.

What is their magic ingredient? In a word, pressure. The authors' experience with computer simulations of liquids akin to silicon and germanium led them to the idea of adding an additional variable with which they could tune a liquid's ability to thwart crystallization. Much as composition can be used in this way for multi-component alloys, so pressure could be the tool for monatomic materials. It is not, however, a universal answer: a necessary condition is that applying pressure reduces the temperature at which the crystal melts, and this occurs only in materials where crystallization takes place with an increase in volume.

This unusual thermodynamic behaviour is familiar from the case of water (liquid water expands as it crystallizes to ice), and also occurs more generally in liquids that form local tetrahedral atomic arrangements at low temperature and low pressure, as do silica, silicon and germanium. And Bhat *et al.* give convincing arguments that, at a range of high pressures, liquid germanium, when cooled quickly, produces a glass. They do not, however, provide an *in situ* characterization of this glassy phase, and more work will clearly be needed to describe the structure and properties of this state fully, and to investigate more extensively whether glass can be formed in related materials such as silicon.

Last but not least, Bhat and colleagues' work¹ addresses another intriguing phenomenon whose existence has been posited in recent years. It seems that the local arrangements of the atoms or molecules of liquids such as water, silica, silicon or germanium are different at low pressures and temperatures from those seen at high pressures and temperatures. In the former, they tend to exist as a fairly open, tetrahedrally bonded network; in the latter, as a more tightly packed structure.

Whether or not these two different local structures amount to two different liquid phases separated by a thermodynamic phase transition has been a point of debate. The authors give strong, but indirect, evidence that such a liquid–liquid transition does occur in supercooled germanium. Understanding the effect of the competition between different local liquid structures and its relation to crystallization and glass formation remains a formidable theoretical challenge. ■

Gilles Tarjus is in the Laboratoire de Physique Théorique de la Matière Condensée

(CNRS-UMR 7600), Université Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris, France.
e-mail: tarjus@lptmc.jussieu.fr

1. Bhat, M. H. *et al.* *Nature* **448**, 787–790 (2007).

2. Turnbull, D. *J. Appl. Phys.* **21**, 1022–1028 (1950).
3. Turnbull, D. *J. Metals* **188**, 1144–1148 (1950).
4. Turnbull, D. *Contemp. Phys.* **10**, 473–488 (1969).
5. Johnson, W. L. *MRS Bull.* **24** (10), 42–56 (1999).
6. Inoue, A. *Acta Mater.* **48**, 279–285 (2000).
7. Franck, F. C. *Proc. R. Soc. A* **215**, 43–46 (1952).

PARKINSON'S DISEASE

Pro-survival effects of PINK1

Asa Abeliovich

However parkinsonism is initiated, the progressive symptoms are similarly devastating. So insights from analyses of gene mutations linked to these disorders should aid a better general understanding of them.

The second most common neurodegenerative disorder, Parkinson's disease, is caused by a loss of dopamine-secreting neurons from the mid-brain region; this leads to rigidity and tremor at rest and the characteristic slowness of movement. What triggers Parkinson's disease and parkinsonism (any of several nervous-system disorders that are similar to Parkinson's disease) remains elusive, but some forms of these disorders are heritable. In a paper in *PLoS Biology*, Pridgeon *et al.*¹ attempt to elucidate the mechanism behind a familial form of Parkinson's disease that is associated with mutations in the *PINK1* gene. Their findings point to an old culprit — the mitochondrion, the cell's energy-producing organelle — and a new offender, the molecular signalling pathway involved in inhibiting programmed cell death (apoptosis).

Parkinson's disease presents considerable challenges to the researcher². First, its insidious progression within the ventral midbrain region makes detailed pathological analysis of the disease onset difficult. Second, no animal model accurately recapitulates the spectrum of features of this disease, whether at the level of the organism or of its cellular hallmarks (such as the loss of dopamine-secreting neurons and the accumulation of proteinaceous inclusions termed Lewy bodies within the cytoplasm). Third, epidemiological evidence suggests a complex multifactorial aetiology involving both environmental and genetic factors.

Nonetheless, there is much optimism about progress in the field as a result of the discovery of a trove of genes associated with parkinsonism that are mutated in the inherited forms of these disorders. Such mutations account for a minority of cases of Parkinson's disease, but it is reasonable to think that similar mechanisms might underlie the inherited and the sporadic forms of the disease.

Although no clear unifying theme has emerged from the analysis of parkinsonism-associated genes, a potential site of action is the mitochondrion. Epidemiological studies have implicated mitochondrial toxins such as pesticides in the aetiology of the disease³. From

correlative studies, the activity of mitochondrial complex 1 (which is involved in the production of the cellular energy source, ATP) is impaired in patients with Parkinson's disease³. Moreover, at least three of the genes associated with the familial form of this disease have been linked to mitochondria. For example, although the products of the parkin and *DJ-1* genes, which are mutated in some forms

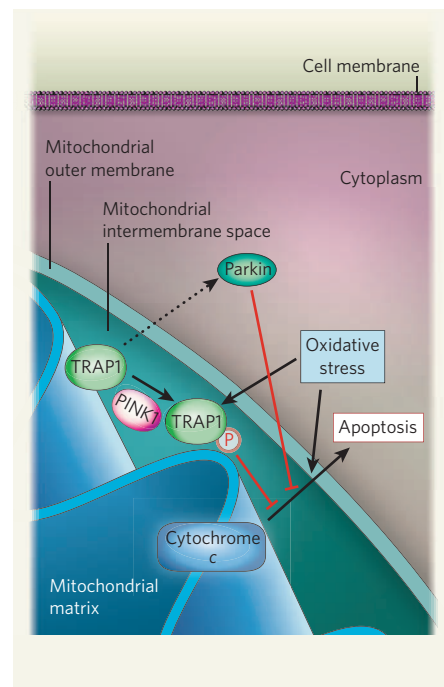


Figure 1 | Parkinsonism and signalling within mitochondria. The work of Pridgeon and colleagues¹ implicates PINK1 in the prevention of parkinsonism-associated cell death, which is thought to be induced by oxidative stress. This enzyme seems to mediate phosphorylation of TRAP1; in turn, TRAP1 prevents the release of cytochrome *c* from the mitochondrial intermembrane space into the cytoplasm, where it is involved in apoptosis. The protein parkin, which has also been linked to inherited forms of parkinsonism, inhibits apoptosis as well, although it remains unclear whether PINK1 and parkin interact directly or through intermediaries.

of parkinsonism, accumulate mainly in the cytoplasm, they are also reported to associate physically with mitochondria^{4,5}. Moreover, mutations in mouse *parkin*, as well as in one of its *Drosophila* homologues, lead to defects in mitochondrial structure and energetics^{6,7}.

The most direct link between mitochondria and a parkinsonism-associated gene comes from the discovery⁸ of *PINK1* mutations in some patients. The *PINK1* protein carries both a putative sequence acting as a localization signal to take it to mitochondria and a kinase domain with the enzymatic activity to phosphorylate serine and threonine amino-acid residues.

So what could be the function of *PINK1* in mitochondria? Although *Pink1*-deficient mice are reported to lack overt mitochondrial defects⁹, mutations in a *Drosophila* homologue of *PINK1* lead to deformed mitochondria, particularly in muscle and gonadal cells^{10,11}. This outcome is similar to that of mutations in the *parkin* gene of *Drosophila*⁷, indicating that these two genes might share a conserved genetic pathway (Fig. 1). Indeed, overexpression of *parkin* can overcome the anomalies associated with *Pink1* mutations in *Drosophila*^{10,11}.

To delve further into the molecular mechanism of *PINK1* action, Pridgeon *et al.*¹ identified one of its binding partners, TRAP1. They found that both *PINK1* and TRAP1 accumulate in the mitochondrial intermembrane space and inner membrane. Moreover, when the authors overexpressed *PINK1* in cells, TRAP1 phosphorylation was enhanced, as it also was in response to oxidative stress. Consistent with earlier reports⁸, the authors found that the overexpression of normal *PINK1*, but not the parkinsonism-associated mutant, protects cells from apoptosis in response to oxidative stress. However, even *PINK1* overexpression could not prevent oxidative-stress-induced apoptosis when TRAP1 levels were reduced. Together, these data indicate that *PINK1* acts upstream of TRAP1 within an anti-apoptotic signalling cascade.

The regulation of apoptosis by TRAP1 as observed by Pridgeon and colleagues seems to be broad-spectrum, because this protein also protects cells against other toxic insults such as the chemotherapeutic agent VP16, or etoposide¹². The exact mechanism by which TRAP1 suppresses apoptosis is unclear; it does not seem to interact physically with cytochrome *c* (ref. 1) — a mitochondrial protein that, when translocated to the cytoplasm, participates in inducing apoptosis.

Other questions also remain. Pridgeon *et al.*¹ performed their studies in cultured cell lines. Whether the *PINK1*-mediated signalling pathway also functions in neurons, and in the context of Parkinson's disease, remains to be seen. Another question is whether *Parkin* is a component of the same pathway.

By implicating *PINK1* in an anti-apoptotic mechanism within mitochondria, the work of

Pridgeon *et al.* hints that this protein — and other components of the signalling pathway in which it functions — is a potential therapeutic target for Parkinson's disease. It remains possible that *PINK1* also functions in other crucial mitochondrial processes, including the regulation of oxidative phosphorylation, intracellular buffering of calcium and the generation of reactive oxygen species, all of which are essential for neuronal survival and might thus be implicated in Parkinson's disease. ■

Asa Abeliovich is at the Taub Institute for the Aging Brain, Departments of Pathology and Neurology, Center for Neurobiology and Behavior, Columbia University, 630 West

168th Street, New York, New York 10032, USA.

e-mail: aa900@columbia.edu

1. Pridgeon, J. W., Olzmann, J. A., Chin, L. S. & Li, L. *PLoS Biol.* **5**, e172 (2007).
2. Abeliovich, A. & Flint Beal, M. *J. Neurochem.* **99**, 1062–1072 (2006).
3. Beal, M. F. *Ann. NY Acad. Sci.* **991**, 120–131 (2003).
4. Darios, F. *et al. Hum. Mol. Genet.* **12**, 517–526 (2003).
5. Zhang, L. *et al. Hum. Mol. Genet.* **14**, 2063–2073 (2005).
6. Palacino, J. J. *et al. J. Biol. Chem.* **279**, 18614–18622 (2004).
7. Greene, J. C. *et al. Proc. Natl Acad. Sci. USA* **100**, 4078–4083 (2003).
8. Valente, E. M. *et al. Science* **304**, 1158–1160 (2004).
9. Kitada, T. *et al. Proc. Natl Acad. Sci. USA* **104**, 11441–11446 (2007).
10. Clark, I. E. *et al. Nature* **441**, 1162–1166 (2006).
11. Park, J. *et al. Nature* **441**, 1157–1161 (2006).
12. Masuda, Y. *et al. J. Biol. Chem.* **279**, 42503–42515 (2004).

ASTROPHYSICS

Photons from a hotter hell

Trevor Weekes

Blazars are massive black holes sending out particle jets at close to the speed of light. Stupendously fast, intense bursts of highly energetic γ -rays indicate that the blazar environment is even more extreme than was thought.

Serendipity has always played a large part in astronomy. Detecting the short-lived, violent phenomena characteristic of high-energy astrophysics is a case in point. Catching these transient signals as they appear, dominate the sky briefly, and disappear again — perhaps never to be repeated — requires not only the right telescope, but also the luck of pointing it in the right direction. When technology and serendipity do come together, dramatic results can follow.

An example of such an auspicious conjunction is given by two papers from the *Astrophysical Journal*^{1,2}, in which two separate teams of astronomers report the detection of powerful bursts of teraelectronvolt (TeV) γ -rays lasting just minutes, the shortest time ever observed. The sources, billions of light years away, are

two 'blazars' — black holes of more than 100 million solar masses that signal their presence through jets of charged particles emitted at almost the speed of light.

The detection of high-energy γ -ray emission from blazars is not new. The γ -ray telescope EGRET, on NASA's Compton γ -Ray Observatory, was sensitive to photons 100 million times more energetic than optical photons, and reported the detection of some 70 blazars³ almost a decade ago. The new generation of telescopes, with acronyms such as CANGAROO-III, HESS, MAGIC and VERITAS, is sensitive to TeV γ -rays 1,000 times more energetic again, and has already catalogued some 60 sources, including 15 blazars^{4,5}. In the Universe that is being revealed by these telescopes, violent, high-energy phenomena are commonplace.

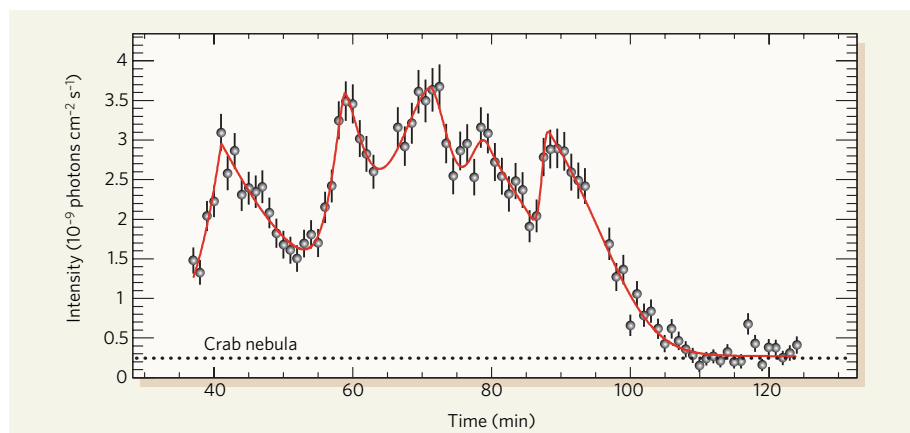


Figure 1 | Cosmic rollercoaster. The γ -ray flux from the blazar PKS 2155–304 at energies above 0.2 TeV, observed by HESS¹ on 28 July 2006. Five overlapping emission peaks were seen, each with rise times of just a few minutes. The data are binned in one-minute intervals; the horizontal line shows the flux from the Crab nebula, the strongest steady source in the TeV sky. (Plot reproduced from ref. 1.)



Figure 2 | Eye on the sky. The 17-metre-aperture MAGIC γ -ray telescope at the Roque de los Muchachos site on La Palma in the Canary Islands.

The new findings^{1,2} are based on the atmospheric Čerenkov technique, in which a γ -ray is detected indirectly through a shower of secondary particles that initiates an optical shock wave as it passes through the atmosphere. The blue light produced in this process can be easily detected by large, relatively crude optical telescopes coupled with fast, sensitive electronic cameras.

The High Energy Stereoscopic System (HESS) is one such observatory, comprising an array of four telescopes of 12-metre aperture in the central Namibian highlands. Currently the most sensitive instrument at TeV energies, HESS has since 2002 been routinely observing the blazar PKS 2155–304, a known emitter of high-energy γ -rays. Demand for observing time on this instrument is such that the variable emission of one blazar cannot be monitored continuously over a long period. But early in July 2006, the HESS observing team noted that this particular blazar was in an unusually high state of emission, and they alerted other observers to initiate a multi-wavelength observing campaign.

In this way, many eyes were fortuitously looking in the right direction when, in the early hours of 28 July, PKS 2155–304 suddenly flared up in a series of five overlapping bursts within one hour (Fig. 1). Some of these flares doubled in intensity in as short a time as 3 minutes, and the peak intensity was some 15 times that of the Crab nebula, the brightest steadily emitting 'standard candle' at TeV energies.

The normal, quiescent brightness of PKS 2155–304 is only 15% that of the Crab nebula, and so this was the strongest outburst ever detected at TeV energies from any source, either within our Galaxy or outside it. The recent paper¹ contains details only of the time

structure and energy spectrum of the burst. It shows, for instance, that the distribution of energies in the spectrum did not vary substantially during the flare. Publication of the rest of the data promises a field day for theorists working on blazar-jet phenomena.

The huge and sudden short-term variability of blazars found here is supported by a set of observations² from another telescope, MAGIC (Major Atmospheric Gamma Imaging Čerenkov). Located in the Canary Islands, MAGIC is the world's largest single atmospheric Čerenkov telescope, with an aperture of 17 metres (Fig. 2). On 30 June and again on 9 July 2005, it saw short flares from the blazar Markarian 501, a well-known and highly variable TeV emitter. These flares had lower peak fluxes than the PKS 2155–304 outbursts, but even shorter intensity doubling times — as little as 2 minutes.

These observations^{1,2} are important because they place new constraints on the dynamics and dimensions of blazar systems. This in turn matters not only because blazars might be the source of ultra-high-energy cosmic rays and more generally of the γ -ray bursts now continually spotted by astronomers, but also because they are the only cosmic laboratories in which extreme physics can be studied.

The accepted view of blazars is that they are a sub-class of active galaxies (formerly loosely termed quasars) consisting of supermassive black holes whose disk of accreting materials results in the formation of relativistic jets of gas at right angles to their plane. In blazars, one of these jets just happens to point in our direction. This outflow accelerates a blob of particles, either protons or electrons, to nearly the speed of light, and these particles in turn emit the γ -rays. One might therefore assume that a



50 YEARS AGO

"Report of the Tobacco Manufacturers' Standing Committee" — The report... shows undisguised attempts to belittle the findings of those investigators who have shown a correlation between smoking and lung cancer... The report brings its heavy guns to bear on the statistical aspect of the problem. It emphasizes that a contingent statistical relation does not guarantee causation. Let us take an example of what could be called contingent. Seaside-sunburn in London school children, before the motor-car, was always preceded by a railway journey. The railway journey is contingent to the sunburn but is not the cause of the sunburn. To take a strictly practical point of view, the seaside-sunburn could have been avoided by shutting down the railways, and in the same way lung cancer could be largely avoided by closing the cigarette factories, quite independently of whether lung cancer and smoking have a causal or a contingent relationship. From *Nature* 17 August 1957.

100 YEARS AGO

"The Second International Congress on School Hygiene" — Dr Schuyten (Antwerp) presented a summary of ten years of research in the paedological laboratories of Antwerp... The chief general conclusions are, (1) that the child, on entering the ordinary school, undergoes physical and mental depression; (2) that growth in muscular power is not regular during the school year, there being a distinct depression in March; (3) that, as tested by the dynamometer, muscular power varies with the season; (4) that voluntary attention decreases from January to July, and increases from October to December; (5) that fatigue increases during the school year from one end to the other without perceptible recovery of energy due to holidays; (6) that the validity of aesthesiometric methods of determining fatigue is now demonstrated. From *Nature* 15 August 1907.

50 & 100 YEARS AGO

flare of 3 minutes' duration indicates an emitting blob 3 light minutes long, which is tiny compared with the size of a galaxy.

This assumption neglects relativistic effects, which stretch the blob and increase its size in proportion to the relativistic Doppler factor. A previous observation of a TeV flare from the blazar Markarian 421 indicated that this factor was around 10 (ref. 6). But the black hole associated with PKS 2155–304 is predicted to have a mass of 1 billion to 2 billion Suns, larger than most active galaxies. The length of its flares¹ combined with a Doppler factor of 10 would give an emitting region of only about a tenth of the Schwarzschild radius, a measure of the size of the black hole. Thus, either the Doppler factor must be around 10 times larger (and so the plasma in the jet must be travelling even faster than assumed), or the emission must come from a very compact region of the jet that is comparable to, or smaller than, the size of the black hole. A similar conclusion applies to the Markarian 501 blazar observed by the MAGIC Collaboration² (although in this case the situation is less dramatic, owing to the smaller assumed size of the associated black hole).

Given that TeV-emitting blazars have been monitored for only a short time, it is most likely that further, perhaps even shorter, flares will be observed. Significant upgrades are under way for HESS and MAGIC to capture such events; VERITAS, an array with comparable sensitivity to the current HESS, has recently come online in Arizona. GLAST, the next-generation γ -ray space observatory, will add significantly to the catalogue of γ -ray-emitting active galaxies and will monitor their long-term behaviour. It is not likely to add significantly to the detection of fast flares, however, because of the limited collection area inherent in satellite telescopes. The investigation of the detailed structure of blazar jets will remain the prerogative of ground-based γ -ray observatories.

When the first detection of TeV γ -rays from an active galaxy was reported⁷, Francis Halzen described the phenomenon as “photons from hell”⁸. The latest observations^{1,2} indicate that the hell required to produce these photons is even smaller, more violent and more difficult to explain than we thought. ■

Trevor Weekes is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA, and a visiting fellow at the School of Physics, University College Dublin, Ireland.
e-mail: tweekes@cfa.harvard.edu

COMPUTATIONAL BIOCHEMISTRY

Models of transition

JoAnne Stubbe

Is it possible to determine the role of an enzyme from its structure? The latest findings suggest that it is, and prove the point by predicting the substrate for an enzyme of unknown function.

Although genomes are vast sources of information, not all of this information is currently understood. About 600 genomes have been completely characterized¹, and have revealed thousands of DNA base sequences that could encode proteins — such sequences are known as open reading frames (ORFs). But half of these ORFs have no assigned function; furthermore, the roles of many genes have been incorrectly assigned (annotated) in genome databases. These errors lead to additional misassignments as new sequences are deposited. Determining the function of a protein is thus a huge challenge that requires creative, multi-disciplinary approaches.

Reporting on page 775 of this issue, Hermann *et al.*² demonstrate an original solution to this problem that combines computational methods with knowledge of enzyme mechanisms and structures. In this way, they identify the function of a previously unassigned ORF and propose that it is part of a metabolic pathway present in several organisms.

Determining an enzyme's role from scratch is almost impossible, but clues may be gleaned from proteins with similar structures that have known functions. Often these proteins have quite different amino-acid sequences, but at certain positions within their structures — those that are essential for catalysis — their sequences are alike. These groups of enzymes are classified into ‘superfamilies’ of proteins that use similar amino-acid residues to catalyse the same chemical transformation (such as forming a carbon–oxygen bond) but as part of a different overall reaction. The substrates for enzymes within a superfamily can have diverse structures³.

Hermann *et al.*² set out to establish the function of protein Tm0936 found in *Thermotoga maritima* bacteria. This protein belongs to the amidohydrolase superfamily (AHS) of enzymes. AHS enzymes share a structural feature known as an $(\beta/\alpha)_8$ -barrel fold and degrade substrates by reacting them with water at an active site that contains one or two metals. These hydrolysis reactions proceed through intermediates that have a tetrahedral arrangement of chemical groups around the reacting carbon atom³ (Fig. 1a). All AHS proteins that have been studied mechanistically contain the same structural feature: a series of histidine amino acids (which bind to the metals in the active site) on specific strands of the barrel. This characteristic motif can be used to identify other AHS members, using

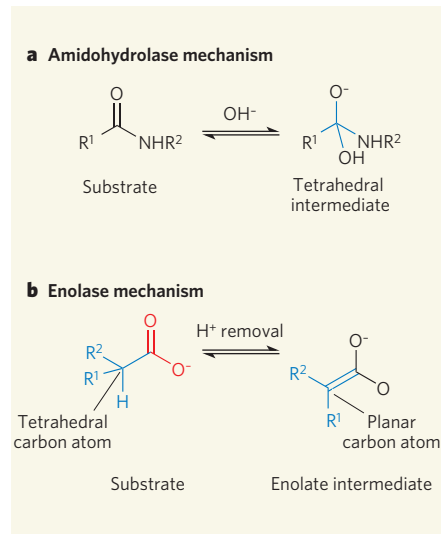


Figure 1 | Chemical transformations of enzyme superfamilies. Enzymes that catalyse the same chemical transformations can be grouped into ‘superfamilies’. **a**, The amidohydrolase superfamily catalyses the degradation of substrates with water, and proceeds through intermediates that have a tetrahedral arrangement of bonds around a carbon atom (blue), as in the example shown. Hermann *et al.*² use this mechanism as the basis of their computational method to predict the substrate of an amidohydrolase protein, Tm0936. **b**, The enolase superfamily catalyses the removal of a hydrogen ion (H^+) from carbon atoms adjacent to a carboxylate group (red) to form enolate intermediates. The arrangement of bonds around the reacting carbon atom changes from a tetrahedral to a planar geometry. R^1 and R^2 represent any hydrocarbon group.

search algorithms that detect distantly related amino-acid sequences in proteins. Knowing that Tm0936 is an AHS enzyme limits the number of possible reactions that it can catalyse, which is a crucial first step in unravelling its function.

To identify the substrate of Tm0936, the authors² started from a central premise of enzyme behaviour — that enzymes bind to the ‘transition state’ of their reaction more tightly than to substrates in their ground states; the transition state is the highest-energy configuration of atoms in the reaction pathway. Unfortunately, transition states cannot be generated computationally because the bonds involved are being made and broken; consequently, the bond lengths and charge distributions are

1. Aharonian, F. A. *et al. Astrophys. J.* **664**, L71–L74 (2007).
2. Albert, J. *et al. Astrophys. J.* (in the press); preprint available at www.arxiv.org/abs/astro-ph/0702008 (2007).
3. Hartman, R. C. *et al. Astrophys. J. Suppl. Ser.* **123**, 79–202 (1999).
4. www.mppmu.mpg.de/~rwagner/sources/index.html (2007).
5. <http://tevcat.uchicago.edu>
6. Gaidos, J. A. *et al. Nature* **383**, 319–320 (1996).
7. Punch, M. *et al. Nature* **358**, 477–478 (1992).
8. Halzen, F. *Nature* **358**, 452–453 (1992).

EARTH SCIENCE

Coastal catastrophe in Phoenicia

The coast of Lebanon — Phoenicia of the ancient world — has seen a lot of action over the centuries, and not just from trading or warring peoples. This is a tectonically complex and dynamic region, and on 9 July, AD 551, a catastrophic earthquake and subsequent tsunami hit cities along some 180 km of coast running from Tripoli in the north to Tyre in the south. Contemporary accounts and archaeological evidence attest to the extent of the devastation.

Ata Elias and colleagues, based at IPG-Paris and Ifremer, have been exploring the geological context of the AD 551 earthquake and tsunami, and that of similar events in the same area. As they describe in *Geology* (A. Elias *et al.* *Geology* **35**, 755–758; 2007), they have come up with a fresh line of evidence for the seismic particulars involved. The cause, they conclude, was

a sudden rupture of the 'Mount Lebanon thrust', a previously unidentified offshore fault about 150 km in length, which runs close to the Lebanese coast. This satellite image shows the region concerned.

The new data come from a geophysical survey involving a towed acoustic system that allowed the deep-sea floor to be imaged. The resulting pictures reveal dramatic submarine scarps and ruptures on the sea floor, akin to those resulting from large-scale seismic activity on land, that have broken through the thick blanket of sediment.

The continental shelf is unusually narrow in this region, and the fault comes as close as 8 km to the coastline. Over geological time, movement on this fault has been responsible for elevation of the Mount Lebanon range.

Evidence for recurrent seismic

activity along this coast comes in particular from sets of raised 'surf-cut benches'. These structures form at sea level; they are marked by the presence of characteristic, reef-like biological constructions that can be dated with carbon isotopes. The lowermost of these benches was violently lifted 80 cm out of the water sometime in the sixth century AD, and probably marks a simultaneous consequence of the AD 551 event that devastated the Phoenician cities.

From similar features, Elias *et al.*

estimate that the Mount Lebanon thrust ruptures with a periodicity of some 1,500–1,750 years. That might mean an earthquake is overdue, but seismic activity is notoriously unpredictable.

Tim Lincoln



NASA/SPL

variable. Hermann and colleagues therefore screened possible substrates for Tm0936 by modelling high-energy intermediates that would be formed during the enzyme reaction. These intermediates closely mimic the transition state for each substrate (Fig. 1a), and so provide better structures than those of the ground-state substrates for modelling binding interactions with the enzyme.

To limit the choice of possible substrates, Hermann *et al.* focused on naturally occurring compounds that are listed in the metabolite database of the Kyoto Encyclopedia of Genes and Genomes¹. Knowing which chemical groups are recognized by AHS enzymes, they examined all of these metabolites and came up with a list of about 4,000 potential substrates. They then modelled tetrahedral intermediates of these compounds, including all the possible stereoisomers that arise in chiral molecules. Using computational methods, the authors docked all of the putative intermediates into a model of the active site of Tm0936 and identified the best-fitting candidates. They then used biochemical assays to test these candidates for catalytic activity and specificity, and identified several compounds that are excellent substrates for Tm0936. With the function of Tm0936 thus revealed, the authors were also able to assign the roles of 78 previously unannotated ORFs of AHS members found in many species. Furthermore, the reaction that Tm0936 catalyses could form part of a previously undiscovered degradation pathway.

Enzyme superfamilies are constantly being

discovered³. Using Hermann and colleagues' approach², it should be possible to determine the functions of members of these families that currently have unknown roles. For example, enzymes of the enolase superfamily (ENS) have a (β/α)₈-barrel structure similar to that of the AHS, and their mechanism has been extensively studied³. They remove a hydrogen ion (H^+) from carbon atoms adjacent to carboxylate groups (CO_2^-) in substrates, generating an intermediate known as an enolate that is stabilized by metal atoms in the enzyme active site (Fig. 1b). This process changes the geometry of the reactive carbon atom from a tetrahedral arrangement of four groups about the carbon in the ground state to a planar arrangement of three groups about that atom in the high-energy intermediate. The planar geometry of the intermediate could be used as the basis of a metabolite search-pool. Chemical expertise is essential for choosing the best bond distances and charge distributions in the computer-generated intermediates, ready for the enzyme-docking analysis.

The AHS and ENS account for a large number of enzymes, so although Hermann and colleagues' method is labour intensive, it may be very useful in assigning enzyme function. Nevertheless, a few problems remain, as the authors have noted². Proteins in the ENS are composed of two domains, one of which does not share amino-acid sequences with other superfamily members, even though it contributes to substrate specificity. Furthermore, multidomain structures often exist in different structural states: open (in the absence

of substrate) and closed (during catalysis; that is, in the presence of a high-energy intermediate). Thus, available crystal structures may not be indicative of the closed state, which would be most relevant to the enzyme's reaction.

Another issue is our limited knowledge of metabolites; more of these must be identified if the authors' approach is to be effective. Finally, Hermann and colleagues' process must be developed into a higher-throughput method if it is to make an impact on this challenging area. Further experiments are required to see if their technique is truly general for all enzymes, but the indications are that it will provide an excellent complement to other methods for assigning roles to unknown ORFs.

JoAnne Stubbe is in the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. e-mail: stubbe@mit.edu

1. www.genomesonline.org
2. Hermann, J. C. *et al.* *Nature* **448**, 775–779 (2007).
3. Gerlt, J. A. & Babbitt, P. C. *Annu. Rev. Biochem.* **70**, 209–246 (2001).
4. Kanehisa, M. & Goto, S. *Nucleic Acids Res.* **28**, 27–30 (2000).

Correction

In the News & Views article "Chemical biology: Ions illuminated" by Christopher J. Chang (*Nature* **448**, 654–655; 2007) an error crept into part a of the accompanying figure. The correct structure of CaGf can be found at <http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?sid=24711513>

OBITUARIES

Anne McLaren (1927–2007)

Inspiring reproductive biologist and mammalian geneticist.

On 6 July, Anne McLaren spent a busy day at the Gurdon Institute in Cambridge, where she had worked since 1992. She prepared a talk for a meeting in Germany and answered a large number of e-mails. In the afternoon, she attended a group leaders' meeting, as always paying close attention and ready to offer sensible advice. Towards the end of the day, she chatted with colleagues and asked questions about some recent stem-cell publications. She left promising to continue the discussion. Sadly, this was to be her last working day.

Anne McLaren had an extraordinary life, both personally and professionally. The daughter of industrialist Henry McLaren, Second Baron Aberconway, and his wife Christabel McNaughten, in 1945 she embarked on the study of zoology at the University of Oxford because for her this was an easier option than reading English, for which the entrance examination required too much reading in too little time. She completed her doctoral studies in 1952, and moved to University College London. There she began her studies on mouse genetics and reproduction with her colleague Donald Michie, whom she married that same year.

Initially, McLaren's research interest was in the interactions between genes and the environment. One of her findings — now often ignored in bioassays and drug testing in mice — demonstrated that, compared with the offspring of a cross-strain mating, inbred strains of mice showed greater variability in their response to stress. These ideas were elegantly recaptured in a review, "Too late for the midwife toad", written more than 40 years later. The article encompasses not only Conrad Waddington's theories of canalization and the inheritance of apparently acquired characteristics, but also the recent molecular explanations for morphological evolution based on studies in flies.

But McLaren's abiding interest in "everything involved with getting from one generation to the next" began with an observation on the differences in the number of lumbar vertebrae in two strains of inbred mice. She wanted to know whether such variability was due to inherent differences between the embryos of the two strains or because of differences in their uterine environments. To answer this question, McLaren and her colleagues induced

ovulation in mice, retrieved fertilized embryos from one strain and transferred them into females of the other. They found that the uterine environment influences the outcome.

This work was the precursor to a collaboration with John Biggers, with whom she showed that early mouse embryos could be cultured for a day or two *in vitro* and go on to develop into adult animals after transplantation into the uteri of surrogate females. This study was to capture the public imagination. And it provided an essential backdrop to reproductive research in humans that led to the development of *in vitro* fertilization.

McLaren continued her flourishing work on reproductive biology and early development at the ARC Unit of Animal Genetics in Edinburgh, and in 1974 she returned to University College London as the director of the newly established MRC Mammalian Development Unit. During this period, she wrote two highly influential books: *Mammalian Chimaeras* (1976) and *Germ Cells and Soma* (1981). She became increasingly interested in germ cells — the cells involved in reproduction — which she described as "the most fascinating cells of all". She was also interested in sex determination, genetic

imprinting and the X chromosome.

McLaren's knowledge and wisdom made her a valuable member of many societies and committees. Of particular significance was her membership of the Warnock Committee, which advised the British parliament on potential developments in reproductive medicine and subsequently led to the 1990 Human Fertilisation and Embryology Act. For ten years, McLaren served with the Human Fertilisation and Embryology Authority, which regulated the practice of human *in vitro* fertilization in Britain, and she continued to participate in many important debates on the ethics of reproductive technologies and stem cells. However, her concerns were not restricted to human welfare — she was also a co-founder of the Frozen Ark Project, which aims to collect the DNA and cells of endangered animals before they become extinct.

McLaren held many prominent offices. She was the first female officer of the Royal Society, serving as its foreign secretary from 1991 to 1996. In this capacity, she travelled extensively to stimulate and promote excellence in science. No matter where she went, she travelled with a single rucksack and a plastic bag full of research papers. Among many other offices, she was president of the British Association for the Advancement of Science, the Association for Women in Science and Engineering, and the British Fertility Society. In all of these capacities, McLaren particularly enjoyed the opportunity to engage with young and aspiring scientists. Indeed, she was an enthusiastic and

popular teacher at the annual Mouse Embryology Course at Cold Spring Harbor Laboratory in New York.

She received many awards and prizes for her work; she was elected to the Royal Society in 1975 and received its Gold Medal in 1990. She was also a joint recipient of the Japan Prize, and of the March of Dimes Prize in Developmental Biology.

Anne McLaren was frugal in her personal life, but displayed great generosity towards those who sought her help; she always had a spare bed for a visitor or student who needed a refuge. She was passionate about social justice, and frequently emphasized that scientific advances should be for the welfare of all. She was a member of the Communist Party of Great Britain during the cold war, was committed to socialism, and enthusiastically participated in anti-war demonstrations. At the ceremony at which she received her joint award of the Japan Prize, Anne chose to hear *Where have all the flowers gone?* sung by Joan Baez as "a lament for all wars", and John Lennon's *Imagine*, which she said is "about a world of



peace and love and social harmony”.

McLaren and Michie had three children together. Although they had divorced in 1959, they remained good friends and started to live together again in 2005. They died together in a car accident on 7 July 2007, while travelling from Cambridge to London. Memorial funds in support of young

scientists have been established in their memories.

Azim Surani and Jim Smith

Azim Surani and Jim Smith are at the Wellcome Trust Cancer Research UK Gurdon Institute, Tennis Court Road, Cambridge CB2 1QN, UK. e-mails: as10021@mole.bio.cam.ac.uk; jim@gurdon.cam.ac.uk

Donald Michie (1923–2007)

Father of artificial intelligence in Britain.

Donald Michie was a man of many parts, irrepressible energy and great personal charisma. His first significant scientific contribution came in the field of genetics. With his second wife Anne McLaren, tragically killed in the same car accident in which he died, he published research in the 1950s that helped lay the groundwork for modern reproductive technology, and later earned him a Pioneer Award, with McLaren, from the International Embryo Transfer Society. But even while developing those ideas, he became inspired by a very different scientific passion: intelligent machines.

He had first encountered this revolutionary idea in the company of Alan Turing at Britain's code-breaking centre, Bletchley Park, during the Second World War. The two did not cooperate closely, for whereas Turing was busy breaking the Enigma codes, Michie focused on using the Colossus computer to crack messages generated by the Germans' 'Tunny' teleprinter machines. They became friends nonetheless, and regularly played chess. They often discussed the possibility of intelligent machines that could play in their stead — and do mathematics, use language, interpret photographic input, learn, and even (Turing suggested) wander around the countryside unaided.

This shared vision of artificial intelligence (AI) became Michie's guiding principle, and by 1948 he was writing a paper-and-pencil chess program in his spare time. As he said years later, recalling how Turing's speculations had gripped him, “I resolved to make artificial intelligence my life as soon as it became feasible”. He himself helped to do that, not only intellectually, but also commercially, writing the first marketable ‘expert-system shell’ for logic programming in the 1960s. In Britain, he founded AI

virtually single-handed.

His first task was to overcome what he later called “the national computer-blindness”. Even Britain's science minister in the early 1960s knew nothing of the wartime code-breaking efforts, and thought that ‘computing’ meant desk calculators.



During his relentless lobbying, Michie persuaded the Royal Society to provide “a few hundred pounds” to enable him, with Bernard Meltzer, to set up a small AI research group at the University of Edinburgh in 1963. Its existence was made official as the Experimental Programming Unit in 1965,

and in 1967 it became the Department of Machine Intelligence and Perception. Right from the start, the department received frequent visits from AI pioneers in the United States based at the Massachusetts Institute of Technology, Carnegie Mellon University and Stanford University. Its successor is still a leading centre in the field.

Those early years were not free of problems. The initial offer of a deconsecrated church as lab space was withdrawn when Michie's intention to build an intelligent robot became known. But he remained indefatigable, and in the ensuing years fought lustily for the infant discipline, confronting academia, research councils and industry. He also did much to raise the profile of AI with the general public, holding popular lectures and penning one-off articles for the press, as well as a regular column for the trade magazine *Computer Weekly*.

Sometimes his optimism and enthusiasm, for instance in describing the achievements and potential of the Edinburgh robot, went too far, and prompted a backlash from sceptics and rivals. One result was the UK Science Research Council's notorious Lighthill report of 1973, which in effect pronounced research in AI to be a waste of

time and money. Its criticisms were so fierce as to cause a scandal. It was the only ‘internal’ research-council report to be published, together with extensive rebuttals — including nine pages from Michie.

The Lighthill report stalled AI research in Britain for a decade. Morale and funding reached a low ebb, with several prominent researchers fleeing to the United States. Michie was also sidelined at Edinburgh: a new Department of Artificial Intelligence was formed and inherited most of its predecessor's resources, whereas Michie was put in charge of an independent Machine Intelligence Research Unit, and forbidden to work on robotics.

Research in AI was officially rehabilitated ten years later, when in 1982 Japan launched its ‘Fifth Generation project’, a huge financial and industrial commitment to base its future economy on AI, predicting world supremacy as a result. The response, in Britain as in the United States, was an injection of government funds into both military and civilian AI research. Michie's urbane comment was that the Lighthill incident had been a “mishap of scientific politics” due to all-too-human frailties — specifically, “nothing but ignorance at the top”.

Michie's contributions to the theory of AI began in the 1960s, when a colleague had insisted that learning machines were impossible, and had challenged Michie to prove him wrong. Edinburgh still lacked a digital computer, and so Michie built “a contraption of matchboxes and glass beads” to master the art of noughts and crosses (tic-tac-toe). He called it the Matchbox Educable Noughts And Crosses Engine, or MENACE. He followed that in 1968 with the first reinforcement-learning program, the ‘pole balancer’, involving a pole balancing on a cart. This was no simulation, but controlled a real pole balanced on a real cart. In the 1970s, he capped this with research into chess endgames.

Michie's pioneering Graph Traverser of the 1960s provided ideas that are now standard in heuristic search algorithms, and live on in widely used AI planning techniques. His last important contribution was his StatLog project of the early 1990s, a highly insightful comparison of various models for learning algorithms, from statistical approaches through symbolic tree-building to dynamical systems.

From his long career, Donald Michie has left a generous academic legacy: the establishment of AI in Britain. That was a product of his mercurial intelligence, but those who knew him personally will also remember his wide learning, his wit and his charm.

Margaret Boden

Margaret Boden is in the Department of Informatics, School of Cognitive and Computing Sciences, University of Sussex, Falmer, Brighton BN1 9QJ, UK. e-mail: m.a.boden@sussex.ac.uk

The common biology of cancer and ageing

Toren Finkel¹, Manuel Serrano² & Maria A. Blasco²

At first glance, cancer and ageing would seem to be unlikely bedfellows. Yet the origins for this improbable union can actually be traced back to a sequence of tragic—and some say unethical—events that unfolded more than half a century ago. Here we review the series of key observations that has led to a complex but growing convergence between our understanding of the biology of ageing and the mechanisms that underlie cancer.

Like so many areas of science, our subject is one that has no true beginning, and as yet, no clear ending. However, if we must begin somewhere, it would be in the winter of 1951, when a 31-yr-old woman and mother of five small children underwent a seemingly routine biopsy for a suspicious cervical mass. A portion of that biopsy went as usual to the pathology lab for diagnosis; unbeknownst to the patient, another portion was diverted to the research laboratory of two investigators at Johns Hopkins, George and Martha Gey. The Geys had spent the better part of the preceding twenty years attempting to find a human cell that could grow indefinitely in laboratory culture conditions. That search would end with the arrival of this particular biopsy sample. Unfortunately for the patient, the pathology laboratory quickly confirmed that the mass was indeed cancer and, despite surgery and radium treatment, the patient succumbed to her disease a mere eight months later. On the day of her death, in October 1951, George Gey appeared on national television in the United States to announce that a new era in medical research had begun. For the first time, he explained, it was now possible to grow human cells continuously in the laboratory. He termed the cell line he had created the 'HeLa cell', in memory of Henrietta Lacks, the unfortunate young mother whose biopsy sample made all this possible.

Over the next 50 years, researchers would slowly strip away many of the secrets of how a cancer cell achieves and maintains its immortality. Here we review those efforts in an attempt to give both a historical perspective and an update on the more recent experimental highlights. In particular, we will focus on five aspects of cancer biology that appear to be particularly informative about normal ageing: the connection between cellular senescence and tumour formation; the common role of genomic instability; the biology of the telomere; the emerging importance of autophagy in both cancer and ageing; and the central roles of mitochondrial metabolism and energetic-dependent signal transduction in both processes. Together, these findings seem to indicate that both cancer and ageing represent complex biological tapestries that are often—but not always—woven by similar molecular threads.

Cellular senescence as a block to tumour formation

The Geys' success in cultivating human cancer cells spurred a huge interest in isolating as many types of human cells as possible. These early 'cell culturists' quickly recognized that few, if any, of the isolated cell lines maintained a diploid status. This problem led Leonard Hayflick and Paul Moorhead to turn their attention to a particular source of tissue that is now off limits to many scientists. Using human fetal explants, these investigators found that it was possible to grow and maintain normal diploid fibroblasts. Hayflick and Moorhead

emphasized that such isolates were not clonal cell lines, but polyclonal pools or strains¹. Despite their success in growing these cell lines for several months, they soon stumbled upon another curious phenomenon: cells could not be subcultivated more than about 50 times. They noted that the culture medium was not to blame because if they took early passage cells and transferred them to the conditioned media from late passage cells "luxurious growth invariably results." On the basis of this and other arguments, such as the fact that frozen cells retained the memory of their subcultivation history, they concluded that some intrinsic factor/s (later termed 'Hayflick factors') accumulated in these cells until they 'senesced'¹. In a further leap of speculation, they proposed that this cellular phenomenon could be relevant for organismal ageing. The degree of that relevance remains hotly debated, although it is now clear that the senescent response can be triggered by a wide variety of cellular stresses including the loss of telomeres, the de-repression of the cyclin-dependent kinase inhibitor 2a (*CDKN2a*, also known as *INK4a* or *ARF*) locus, or the accumulation of DNA damage and the subsequent activation of the DNA damage response. Furthermore, the critical executioners of senescence in response to the above factors seem to include the well known tumour suppressor pathways that are controlled by retinoblastoma 1 (RB1) and P53, proteins that have been widely implicated in tumorigenesis.

The *CDKN2a* locus encodes two separate proteins including p16INK4a, which regulates RB1 activity by directly inhibiting the cyclin-dependent kinases, and p19ARF, which regulates the function of P53. Soon after the initial discovery of p16INK4a (ref. 2) it was observed that the levels of p16INK4a increased progressively with the proliferative history of cells in culture^{3,4}. Moreover, it was found that the expression of certain oncogenes accelerated the de-repression of the *CDKN2a* locus, a phenomenon that was called 'oncogene-induced senescence'^{5,6}. Since then, a wealth of data has supported a model in which cells 'memorize' excessive mitogenic stimulation and cell divisions, in part through the de-repression of the *CDKN2a* locus. The mechanistic details of how this happens are starting to emerge, and seem in part to involve the progressive loss of repressive Polycomb complexes⁷. Genes of the Polycomb family encode highly conserved but poorly understood proteins that seem to epigenetically repress many target genes by regulating the level of histone methylation.

The relevance of senescence for cancer protection can be rationalized if senescence is considered as a stress-induced barrier that limits the proliferative potential of damaged cells. In keeping with this notion, recent data have shown that there are abundant senescent cells within tumours, thus moving these observations from the realm of the plastic plate into the arena of real cancer biology^{5,6} (Fig. 1). This

¹Cardiology Branch, NIH, NHLBI, Building 10/CRC 5-3330, 10 Center Drive, Bethesda, Maryland 20892, USA. ²Spanish National Cancer Research Center (CNIO) 3 Melchor Fernandez Almagro Street, Madrid 28029, Spain.

intra-tumoral senescence is thought to be triggered mainly by oncogenic signals that can function in part by de-repressing the *CDKN2a* locus. Another crucial mediator of oncogene-induced senescence seems to be the activation of the DNA damage response (DDR) pathway, presumably by the hyper-replication that is characteristic of cancer cells^{8,9}. With regards to *in vivo* senescence, the hierarchy and crosstalk between induction of the *CDKN2a* locus and the activation of the DDR are not well understood. Another recent surprise was the rapid *in vivo* clearance of tumour cells that undergo P53-triggered senescence^{10,11}. At least in one system, this clearance seems to occur through the activation of the innate immune system¹⁰. This is relevant because standard chemotherapy and radiotherapy might function in part by inducing senescence within the tumour mass¹². In addition, other evidence indicates that the immune system's recognition of tumour cells might require the continuous activation of the DDR pathway¹³. Although they are intriguing, such observations need to be placed in the appropriate context, as it is clear that most human malignancies develop in the presence of a functional immune system. Similarly, tissues that undergo normal ageing can accumulate significant numbers of senescent cells, seemingly without provoking a robust immunological response^{14,15}.

The *CDKN2a* locus is normally expressed at very low levels in most tissues of young organisms¹⁶. It is well established that the *CDKN2a* locus is activated during organismal ageing in both rodents and humans, and the levels of p16INK4a constitute an impressively good overall biomarker of ageing^{16,17}. Interestingly, strategies such as caloric restriction, which extend lifespan, also seem to reverse the age-dependent increase in p16INK4a transcription^{16,18}. The suggestion that p16INK4a is not just a marker but a true effector of ageing remains unproven; however, two independent lines of evidence seem to support this idea. First, a recent spurt of human genetic data indicates that single nucleotide polymorphisms (SNPs) near *CDKN2a* might be associated with ageing and age-related pathologies. Indeed, variants within or near *CDKN2a* were initially associated with overall physical frailty¹⁹ and, more recently, with certain known age-related conditions such as the risk of myocardial infarction^{20,21}. A second line of evidence has centred on the connection between p16INK4a and stem cell biology. Such a link was suspected on the basis of the analysis of mice that are deficient in certain Polycomb genes, which normally repress p16INK4a. These

Polycomb-deficient animals seem to have severe defects in stem cell self-renewal²². Interestingly, age-induced expression of p16INK4a in adult stem cells seems to be associated with widespread impairment in tissue regeneration^{23–25}. Further reinforcing this link, mice that lack p16INK4a have increased regenerative potential in diverse niches including the nervous system, the pancreas and the haematopoietic system. These observations point to p16INK4a both serving as a brake for the proliferation of cancer cells, and also limiting the long-term renewal of stem cells. As such, excessive inhibition of pathways that are linked to cancer might reduce the robustness of various stem cell niches. In essence, cancer prevention might come at the expense of an accelerated decline in tissue regeneration and repair. Although such notions might seem depressing, as discussed below, it is important to note that not every strategy that could protect us from cancer comes at the expense of accelerating ageing.

Genomic instability links cancer and ageing

The maintenance of DNA represents a fundamental and continuous challenge to every cell. Although a full description of the mechanisms underlying DNA repair and genomic stability are outside the scope of this review, suffice it to say that there exist multiple distinct pathways to sense and repair damaged DNA, depending on the nature of the damage (including, for instance, whether one or both strands of DNA are involved) and on the phase of the cell cycle in which the damage occurs. Given that genomic instability is a hallmark of most cancers, it is not too surprising that many of the factors that have been implicated in sensing and responding to DNA damage are altered in human tumours. Perhaps less well appreciated is that genomic instability is also a hallmark of ageing. For instance, in budding yeast, the mother cell can give rise to a smaller daughter cell a finite number of times, termed the replicative lifespan. Interestingly, as the mother yeast cell ages, there is a striking increase in genomic instability, as shown by marked loss of heterozygosity (LOH) in the daughter cell²⁶. A similar age-dependent increase in chromosomal instability has been known to occur in mammals for many years²⁷. Recent evidence indicates that the age-dependent accumulation of somatic mutations might vary significantly between different tissues of the same organism and that these genetic alterations might contribute to the stochastic variation in gene expression that is often seen in mammalian ageing²⁸.

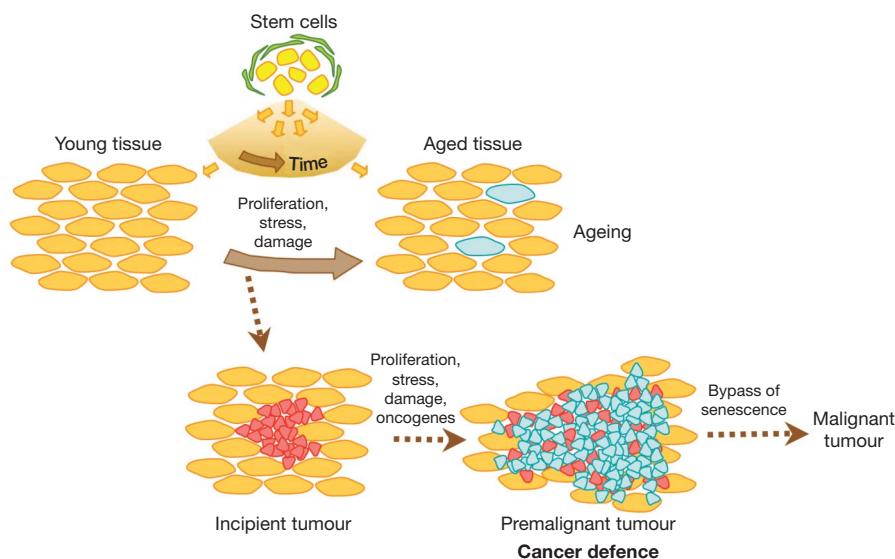


Figure 1 | The potential interplay between stem cells, stress, ageing and cancer. During normal ageing, stem cells accumulate damage and subsequent stress-dependent changes (for example, de-repression of the *CDKN2a* (*INK4a/ARF*) locus or telomere shortening). This leads to the increasing abundance of senescent cells (large blue cells) within differentiated tissues. Incipient tumours, arising directly from stem cells or

from more committed cells, undergo rapid proliferation (small red cells). These pre-malignant tumour cells rapidly accumulate damage, in part owing to the presence of oncogenes, leading to a higher proportion of tumour cells becoming senescent (small blue cells). Tumour progression to full malignancy is favoured when tumour cells acquire mutations that impair the senescence program (for example, mutations in *Trp53* or *CDKN2a*).

The molecular analysis of human inherited cancer syndromes such as Li-Fraumeni syndrome, ataxia-telangiectasia (AT) and common forms of familial breast and ovarian cancer have strengthened the link between the maintenance of genome integrity and cancer susceptibility. These conditions can be caused by germline mutations in the genes for p53, the ataxia-telangiectasia mutated (ATM) kinase and breast cancer 1 (BRCA1), respectively—three proteins that are essential in the surveillance of DNA damage. Interestingly, all three proteins have also been linked to cellular or organismal ageing. In the case of p53, mice that are genetically engineered to express altered isoforms of p53 with increased activity seem to be resistant to cancer but age prematurely^{29,30}. As in all models of accelerated ageing, these results are potentially instructive but extension of their lessons to normal ageing must be done with care. Indeed, this accelerated ageing phenotype might be unique to expression of these altered forms of 'superactive' p53, as simple transgenic overexpression of full-length p53 using its endogenous promoter to regulate expression does not result in a change in lifespan³¹. It is conceivable that normally regulated p53 could have a beneficial impact on longevity by eliminating DNA damage (or DNA-damaged cells). In support of this notion, mice in which both p53 and p19Arf are overexpressed show delayed ageing³², and recent work in *Caenorhabditis elegans* indicates that many lifespan-extending mutations converge on the activation of the worm form of p53 (ref. 33). In a similar vein, a role for the ATM kinase in cellular lifespan was recently established when it was noted that mice lacking *Atm* show rapid ageing and depletion of their haematopoietic stem cell compartment³⁴. Finally, although mice with deletion in *Brcal* die as embryos, mice that express two copies of a hypomorphic *Brcal* allele along with only one copy of the gene for p53 (*Trp53*^{+/-}) survive but show marked progeroid symptoms³⁵. In each of these cases, ageing is the apparent result of either a chronic increase in the level of DNA damage (ATM and BRCA1 deficiency) or the presumed chronic engagement of DNA damage signalling ('superactive' p53).

Recently, the analysis of mice with targeted deletions or humans with inherited deficiencies in various factors that are involved in sensing or repairing DNA damage have significantly strengthened the correlation between DNA damage and the rate of ageing. For instance, reducing the level of various mitotic checkpoint genes leads to chromosomal instability, augmented aneuploidy and the phenotypic appearance of progeroid mice^{36,37}. Similarly, mice lacking the DNA helicase excision repair cross-complementing rodent repair deficiency, complementation group 2 (*Ercc2*, also known as *Xpd*), an enzyme that participates in nucleotide excision repair (NER), show aspects of accelerated ageing³⁸. In addition, a recent report of a single patient with a severe mutation in another NER enzyme, *ERCC4* (also known as *XPF*), showed that this patient also exhibited progeroid features³⁹. Milder mutations in *ERCC4* were previously known to cause the cancer predisposition syndrome xeroderma pigmentosum, a condition that produces a substantial increase in the propensity for various solid tumours. Together these observations indicate that severe deficiencies in proteins that are involved in DNA damage sensing and DNA repair might accelerate ageing, while milder mutations in these same pathways might predispose individuals to cancer. The mechanism for ageing in the absence of faithful DNA repair is not entirely clear but might be secondary to the induction of senescence or apoptosis of crucial stem and progenitor cells. Alternatively, the absence of DNA repair might cause enough direct damage to fully differentiated cells to impair overall homeostasis. By contrast, milder mutations in the same set of genes might allow cells to survive and proliferate, but with an underlying impairment in DNA fidelity that ultimately predisposes to cancer. This may help us to understand why so many mouse models with deletions of genes linked to repairing or sensing DNA damage can exhibit both cancer susceptibility and progeroid features⁴⁰.

Up to now we have discussed a set of genes that were intensively studied because of a known role in cancer or because they were

directly implicated in maintaining genomic stability, and which have been linked to ageing much more recently. Nonetheless, those investigators whose primary focus was studying the molecular basis of ageing have also found a strong reason to focus on genomic stability. This interest in large part came about by pursuing curious observations made more than 100 years ago by two young physicians. In 1904, Otto Werner described four siblings, all of whom had an inexplicable set of symptoms including premature greying of the hair, ageing of the skin, growth retardation and the loss of subcutaneous fat. In the same year, Hastings Gilford coined the term 'progeria' to describe a single case of a teenage boy whose outward appearance suggested that he was considerably older than his stated age of fourteen. A century later, the gene responsible for Werner syndrome was identified and demonstrated to belong to the RecQ helicase family. This family of enzymes is conserved from *Escherichia coli* to man, and humans have five RecQ homologues that have important roles in recombination, the replication stress response and maintaining genomic stability. Given that the gene responsible for Werner syndrome (*WRN*) participates in multiple aspects of DNA homeostasis, it is unclear precisely how the absence of *WRN* activity in affected individuals results in the clinical progeroid syndrome. One attractive explanation is that *WRN* is essential for proper telomere maintenance through its interaction with the telomeric binding proteins TERF1 and TERF2⁴¹. Recent information indicates that the *WRN* gene is often inactivated by epigenetic means in human cancers⁴², indicating that genes that were identified initially as regulating ageing might also inform us about cancer biology. Similarly, the gene responsible for the progeroid Hutchinson-Gilford syndrome has also been identified, and there is evidence that its product helps to maintain nuclear architecture and to regulate genomic stability⁴³.

Finally, another family of genes that has been intensely studied for its role in ageing seems also to have an important function in maintaining genomic stability. This family of proteins is termed the sirtuins, a named based on the family's founding member, the yeast protein silent information regulator 2 (*Sir2*). In yeast, *Sir2* has been implicated in the increase in lifespan that is seen after caloric restriction and in both yeast and worms, overexpression of *Sir2* is sufficient to extend lifespan. One common feature of the sirtuins is their enzymatic function as NAD-dependent deacetylases and for yeast *Sir2*, this biochemical activity is required for the proteins' ability to regulate silencing, recombination and genomic stability⁴⁴. Seven mammalian sirtuins have been identified, with the closest mammalian homologue to yeast *Sir2* being *Sirt1*. Establishing a role of *Sirt1* in epigenetic silencing and genomic stability has been challenging, because mice with targeted deletion of *Sirt1* usually die shortly after birth. Nonetheless, evidence from cell culture models indicates that *Sirt1* is involved in heterochromatin formation and might be particularly important in the silencing of certain tumour suppressors^{45,46}. Interestingly, mice with a targeted deletion in *Sirt6* manifest an accelerated ageing phenotype that on a cellular level is accompanied by genomic instability and defects in DNA repair⁴⁷. It therefore seems that certain members of the mammalian sirtuin family are essential for maintaining genomic integrity, and one could speculate that their absence might eventually be shown to predispose to tumour formation. Curiously, *Sirt6*^{-/-} mice also have severely reduced levels of circulating insulin-like growth factor 1 (IGF-1). Although currently unexplained, this result is interesting because insulin-IGF signalling is known to regulate lifespan in lower organisms. The *Sirt6*-deficient mice therefore add to a growing but poorly understood connection between genomic instability and altered IGF-1 signalling, as insulin-IGF signalling is also perturbed in other mouse models of accelerated ageing, including animals with increased p53 activity³⁰ and mice lacking the NER enzyme *ERCC4*³⁹.

Telomeres and telomerase

Telomeres are specialized structures at the ends of chromosomes, which were first discovered by Barbara McClintock and Herman

Müller in the 1930s on the basis of their ability to distinguish natural chromosome ends from broken chromosomes. Müller first termed these structures telomeres (from the Greek word for 'end part'). Telomeres protect natural chromosome ends from fusion events and are therefore essential for chromosomal stability. However, it was not until after the discovery of the structure of DNA that James Watson and Alexey Olovnikov realized that the replication of the very ends of chromosomes, or telomeres, would probably be impaired because conventional DNA polymerases needed a primer to initiate DNA synthesis. This problem came to be known as the 'end-replication problem' and its solution—or lack thereof—provides yet another link between cancer and ageing.

Molecular insight into the nature of telomeres was not gained until they were first sequenced in *Tetrahymena thermophila* by Joe Gall and Elizabeth Blackburn⁴⁸. Telomeres were found to consist of tandem G-rich repeats that expanded until the chromosome end. Cloning of additional telomeres revealed that they are highly conserved in most eukaryotic organisms, with the exception of some flies. The identification of telomeres rapidly led to the realization that some cells could elongate pre-existing telomeres, implying the existence of telomere elongation mechanisms⁴⁹. This telomere elongation was proposed to rely either on recombination events between telomeres or on the existence of a novel enzymatic activity that could synthesize telomere repeats *de novo*. Such an enzymatic activity was soon discovered by Carol Greider and Elizabeth Blackburn and was named telomerase⁵⁰. A few years later, recombination mechanisms were also shown to maintain and elongate telomeres in the absence of telomerase, the so-called alternative lengthening of telomeres (ALT) pathway^{51,52}.

Most adult cells have limiting amounts of telomerase that are not sufficient to prevent telomere loss, resulting in the shortening of telomeres with age⁵³. As telomeres are essential for chromosomal stability, this progressive telomere loss was proposed to be at the basis of cellular senescence and ageing in the 'telomere hypothesis'⁵³. Indeed, telomere length has been shown to predict the replicative capacity of human cells and the appearance of certain age-associated pathologies in humans^{54,55}. The final demonstration that telomere shortening was one of the causes of cellular senescence came from the observation that re-introduction of the *TERT* telomerase gene was sufficient to bypass replicative senescence and to confer immortal growth on a number of human primary cell lines⁵⁶. Furthermore, the generation and characterization of the telomerase knockout mouse model has been instrumental in demonstrating that short telomeres result in multiple organismal defects caused by defective tissue regeneration^{57–59}. In particular, telomerase-deficient mice with short telomeres show reduced function of various stem cell compartments including those in the bone marrow and skin^{58,60}. Interestingly, patients who have inherited or acquired genetic defects that limit telomere maintenance seem to be at substantially increased risk of a range of conditions including aplastic anaemia⁶¹, idiopathic pulmonary fibrosis^{62,63} and the rarer dyskeratosis congenita syndrome⁶⁴ (Fig. 2).

It remains unclear whether increased telomerase activity can delay or prevent ageing phenotypes in the context of the whole organism. First generation mice lacking the telomerase RNA component *Terc* show a decrease in both their median and maximum lifespan, and these effects become more pronounced with each subsequent generation⁶⁵. The increase in degenerative defects with each subsequent generation that is seen in the *Terc*-deficient mice is known as 'disease anticipation'. This phenomenon is also observed in some patients with dyskeratosis congenita, and is likely to be seen in patients with aplastic anaemia or pulmonary fibrosis due to telomerase deficiencies. As such, ageing or organ failure in these patients might ultimately result from the inheritance of both impaired telomere maintenance machinery and initially shortened telomeres. Conversely, these findings also support the fascinating idea that the manipulation of telomerase activity might increase the life span of

mammals. One caveat to such an approach is the observation that increased telomerase activity in two independent *Tert* transgenic mouse models seemed to increase the susceptibility for tumour formation^{66,67}. In spite of their increased mortality due to cancer, these transgenic mice did show evidence of improved tissue regeneration as well as a slight increase in maximum life span⁶⁸. Interestingly, *Tert* overexpression can also improve the ability of epidermal stem cells to regenerate the skin and the hair^{60,69}.

Over the last decade it has become clear that most human cancers activate telomerase at some point during tumorigenesis, while this activity is largely absent in most normal tissues^{70,71}. A significant number of human tumours can also maintain telomeres by recombination-based ALT mechanisms in the absence of telomerase⁷². By activating a program of telomere maintenance, tumour cells can escape from replicative senescence; this ability was undoubtedly essential for establishing the original HeLa cell line and for most, if not all, immortalized cells thereafter. Conversely, mice with short telomeres are resistant—with a few exceptions—to tumours⁷³, arguing that telomere shortening represents a potent *in vivo* tumour suppressor mechanism.

Autophagy and the cell biology of waste management

The ultimate disposition of cellular waste provides another area in which the biologies of cancer and ageing have merged. One particular area of overlap centres on a process known as autophagy. Originally described in yeast, macroautophagy is the process whereby old and damaged proteins and organelles, including mitochondria, are sequestered into double-membraned structures known as autophagosomes. The autophagosome can then fuse with the lysosome in mammalian cells—or with the vacuole in yeast—to further degrade

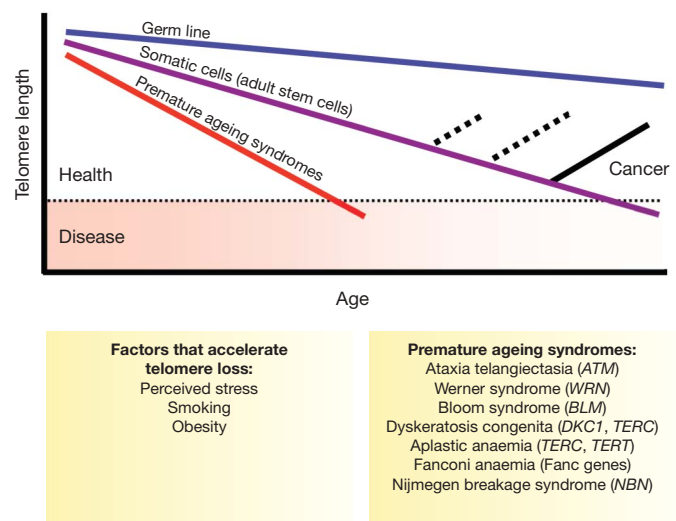


Figure 2 | Revisiting the telomere hypothesis: role of telomeres in cancer and ageing. Normal somatic cells, including adult stem cells, suffer progressive telomere attrition coupled to cell division or to increasing age of the organism. This attrition has been proposed to contribute to multiple age-related pathologies. In germline cells, telomere shortening is attenuated owing to high levels of telomerase activity. By contrast, telomere shortening is accelerated in several human premature ageing syndromes, and patients with dyskeratosis congenita and aplastic anaemia show decreased telomerase activity and shortened telomeres owing to mutations in the *TERC* and *TERT* telomerase genes. Psychosocial and environmental factors such as perceived stress, social status, smoking and obesity have also been shown to accelerate telomere attrition. In contrast to normal somatic cells, most immortalized cultures cell lines and more than 95% of human tumours aberrantly activate telomerase to achieve immortal growth. Although telomerase activity has been shown to be rate-limiting for mouse ageing and lifespan, it is unknown whether increased telomerase activity will be able to extend the lifespan of organisms.

the used cargo back to reusable building blocks. Genetic screens in yeast have identified more than 16 separate and well conserved genes that are required for autophagy to proceed⁷⁴. Relatively little is known about what regulates the rate and selectivity of autophagy. One universal activator of autophagy is a decrease in nutrient availability and this stimulation occurs, at least in part, by the inhibition of TOR (target of rapamycin) signalling⁷⁴.

Although the accumulation of damaged proteins and organelles is a hallmark of ageing and age-related diseases, the link between this process and cancer was, until recently, less clear. The first link between autophagy and cancer came with the observation that the product of the mammalian gene beclin 1 (*Becn1*), a homologue of the yeast autophagy gene *VPS30* (also known as *ATG6*), bound to the human oncogene B-cell CLL/lymphoma 2 (*BCL2*)⁷⁵. Furthermore, the binding of *BCL2* to beclin seems to inhibit the autophagy program. This raised the interesting possibility that, in tumours characterized by increased *BCL2* expression, the oncogene was working at least in part by suppressing autophagy. Further support of this notion came from mice engineered to have lost one copy of the *Becn1* gene. In this case, the haploinsufficient mice developed tumours, indicating that autophagy might act as an important *in vivo* tumour suppressor^{76,77}. This notion is also supported by observations that in human malignancies the *BECN1* locus is commonly deleted and by experimental evidence that several tumour suppressors, including the phosphatase and tensin homologue Pten and p53, stimulate autophagy⁷⁸. Indeed, although the available evidence indicates that oncogenes can inhibit autophagy and tumour suppressors can stimulate it, the exact mechanistic link between cancer and autophagy remains obscure (Fig. 3). One possibility is that interfering with the normal degradation of organelles leads to the retention of older, damaged mitochondria that serve as a source for damaging reactive oxygen species. Such a mechanism would provide a strong connection between common insults that could both contribute to cancer and accelerate ageing.

Recent experiments using mouse knockouts of various autophagy genes have provided considerable insight into the normal physiological role of autophagy, and indicate that this process contributes

to certain age-related pathologies. Mice lacking autophagy-related gene 5 (*Atg5*) are born normally but subsequently die within the first 24 h of life⁷⁹. This death is a result of the inability of these mice to activate the macroautophagy program that is required for energy homeostasis immediately after birth. The short lifespan of these mice obviously precludes analysis of whether they also show accelerated ageing or a cancer predisposition phenotype. Nonetheless, two subsequent studies that used brain-specific conditional knockouts of *Atg5* and *Atg7* revealed that, in these animals, the absence of autophagy leads to a shortened lifespan and an accelerated form of age-related neuronal degeneration^{80,81}. Perhaps the strongest link between autophagy and ageing comes from experiments performed in *C. elegans*. In this organism, loss-of-function mutations in the *daf-2* locus extend lifespan. In the worm, *daf-2* is an insulin/IGF receptor that regulates not only lifespan but also entry into the alternative developmental pathway known as dauer. When developing worms are faced with unfavourable environmental conditions such as limited nutrients, they can enter a state akin to suspended hibernation known as the dauer diapause. Indeed, the environmental conditions that stimulate dauer entry and the conditions that stimulate autophagy are outwardly similar. Interestingly, both dauer formation and lifespan extension by *daf-2* have been shown to require the worm orthologue of beclin⁸². Although these results are clearly informative, the role of autophagy in regulating lifespan in purely postmitotic organisms such as *C. elegans* might be considerably more important than in more complex, renewable mammalian systems where there is normally constant cellular proliferation and cell turnover.

Metabolism links cancer and ageing

In relatively simple organisms such as *C. elegans*, mutations that prolong lifespan are often intimately connected with the ability of the organism to withstand stress, particularly oxidative and metabolic stress. This same stress resistance might also be important to a rapidly growing tumour cell, where the supply and availability of nutrients and oxygen are often precarious. This strategic metabolic overlap has been made more concrete by observations of specific genes that link together the triad of lifespan, cancer and energetics. One such gene is *Trp53*, which encodes p53. We have mentioned that this tumour suppressor is one of the most frequently mutated genes in human cancers and that there is evidence that increased p53 activity can accelerate ageing. Nonetheless, there is a growing link between p53 and cellular metabolism. This link has been strengthened by recent reports that p53 regulates the transcription of two proteins, TP53-induced glycolysis and apoptosis regulator (TIGAR) and the SCO cytochrome oxidase deficient homologue 2 (SCO2), which have key roles in the utilization of glucose and mitochondrial respiration, respectively^{83,84}. Another pathway that allows cells and organisms to adapt to changes in nutrient availability is the TOR signalling network. This pathway has been the subject of many reviews^{85,86}. For our purpose, suffice it to say that TOR is activated in the presence of abundant nutrients, and inactivated under starvation. A number of the upstream regulators of TOR including PTEN, tuberous sclerosis 2 (TSC2), v-akt murine thymoma viral oncogene homologue 1 (AKT-1) and serine/threonine kinase 11 (STK11, also known as LKB1) are frequently altered in human tumours. Similarly, the use of the TOR inhibitor rapamycin is currently being actively pursued as a treatment for human malignancies⁸⁵. The TOR pathway has also received renewed interest for its role in ageing. In many organisms, decreased TOR signalling is associated with the extension of life span^{87–89}. Similarly, under some conditions, TOR signalling seems to be required for the longevity benefits of caloric restriction in yeast⁸⁹. Similarly, in mammalian cells, mammalian TOR (mTOR) seems to be an important regulator of overall mitochondrial metabolism⁹⁰.

The family of forkhead transcription factors represents yet another pathway in which cancer, ageing and metabolism intersect. We have

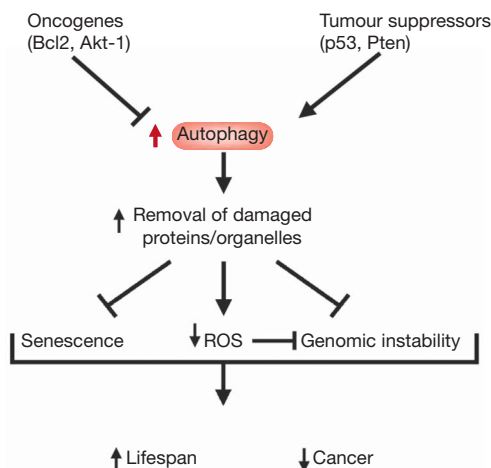


Figure 3 | The potential role of autophagy in cancer and ageing. Autophagy is a regulated process for the removal of damaged proteins and organelles. Autophagy occurs under basal conditions and is stimulated by environmental factors such as starvation. There is evidence that proteins that are linked to tumorigenesis can regulate the rate of autophagy, with oncogenes in general blocking and tumour suppressors stimulating the process. The removal of damaged cellular components, especially damaged mitochondria, might decrease the level of reactive oxygen species (ROS), which in turn might reduce genomic instability or forestall cellular senescence. Such mechanisms might allow moderate increases in autophagy to reduce the incidence of cancer and prolong lifespan.

previously discussed the long lived *daf-2* mutants in *C. elegans*. Characterization of this pathway revealed that longevity mutations in the *daf-2* pathway ultimately function by increasing the activity of the forkhead transcription factor *daf-16*. Interestingly, the transcriptional targets of *daf-16* include a number of enzymes that are involved in metabolism as well as stress resistance⁹¹. Regulation of *daf-16* is in part determined by the subcellular location of the protein and, in worms, the translocation of *daf-16* from the cytosol to the nucleus depends on the availability of nutrients. In mammals, the forkhead family members also regulate numerous aspects of cell fate. The closest mammalian orthologue of *daf-16* is Foxo3, and there is evidence that this mammalian transcription factor also controls the expression of multiple genes involved in stress resistance and metabolism⁹¹. There is also growing evidence that mammalian forkhead family members act as tumour suppressors and regulators of mammalian stem cell lifespan^{92–94}.

Analysis of a very rare tumour that occurs in the carotid body, called a paraganglioma, has led to another interesting link between metabolism, cancer and ageing. The carotid body is located in the neck and serves to stimulate breathing when there is an increase in blood levels of CO₂, or when there is a decrease in O₂ and/or pH. In one large Dutch family with an inherited paraganglioma syndrome, linkage studies revealed a mutation in one allele of the nuclear gene that encodes subunit D of succinate dehydrogenase (SDH)⁹⁵. Subsequent analysis showed that similar SDH mutations could be found in certain pheochromocytomas, a related tumour in the adrenal gland. In addition, analysis of other unrelated families with paragangliomas showed that in these individuals, tumours were associated with mutations in other SDH subunits. SDH is involved in two aspects of metabolism: it functions as an enzyme in the Krebs cycle (converting succinate to fumarate) as well as serving as complex II of the electron transport chain. Interestingly, inherited mutations in fumarate hydratase, another Krebs cycle enzyme, have also been linked to a separate and equally rare inherited cancer syndrome⁹⁶. In both of these cases, tumours seem to arise secondary to an increase in metabolic intermediates (succinate and fumarate, respectively). These intermediates seem to function, in part, by

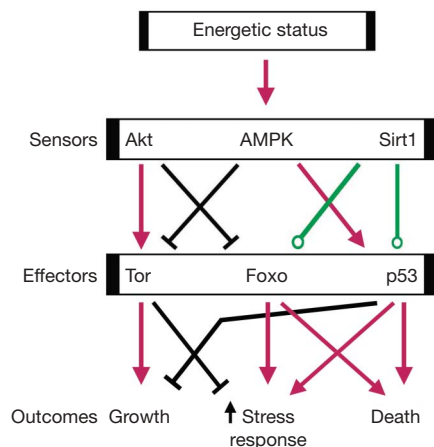


Figure 4 | Energy signal transduction. A complicated web of interactions exists for proteins involved in coordinating energy status with the cell's ultimate fate. Included among these are proteins whose activities seem to be influenced by energetic stores including Akt, AMP-activating protein kinase (AMPK) and the NAD-dependent deacetylase Sirt1. In turn, these sensors can regulate the activity of downstream effectors such as the target of rapamycin (TOR) serine/threonine kinase, the Foxo family of transcription factors and the tumour suppressor p53. Regulation can be positive (red arrows), negative (black lines) or a more subtle change in which some but not all activities are altered (green lines). Overall lifespan seems to be influenced by many of these sensors and effectors, indicating that there is an intimate connection between energy sensing and longevity¹⁰¹.

directly inhibiting the prolyl hydroxylase enzyme family, which controls the degradation of hypoxia inducible factor 1 α (HIF1 α). Interestingly, in lower organisms such as *C. elegans* and *Drosophila melanogaster*, mutations within the same SDH complex can produce an accelerated ageing phenotype^{97,98}. By contrast, knockdown experiments that target other components of the electron transport chain, including essential elements of complexes I, III and IV, lead to lifespan extension in worms^{99,100}. It is unclear why certain mutations in the electron transport chain shorten the life span while others seem to increase it. An attractive but unproven hypothesis is that accelerated or delayed ageing correlates with the level of mitochondria reactive oxygen species that are produced, based on both where and how (for example, knockdown versus structural mutation) the electron transport chain is inactivated.

Finally, we have for the sake of clarity talked about these various pathways in a linear fashion. There is a growing awareness that, as far as the cell is concerned, sensing nutrient availability and energetic status is an integrated and 'system biology' endeavour¹⁰¹. There are many connections between p53, mTOR, Foxo proteins and HIF1- α . For instance, several of these proteins interact directly with each other, and others either share common upstream activators or

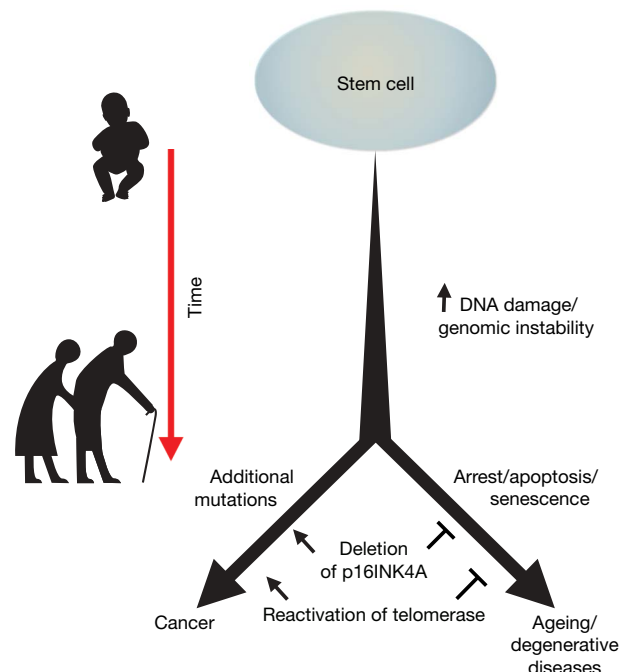


Figure 5 | A stem cell perspective on cancer and ageing. A simplified model that views ageing and cancer from the perspective of alterations within the stem and progenitor cell pool. Over the lifespan of an organism, long-lived cells (such as stem cells) accumulate DNA damage from a number of stresses including intracellular oxidants generated from normal metabolism. The default pathway for such damaged stem cells is to undergo growth arrest, apoptosis or senescence. As more and more stem cells withdraw from the proliferative pool, there is a decrease in the overall number and/or functionality of both stem and progenitor cells. This decrease predisposes the organism to impaired tissue homeostasis and regenerative capacity and could contribute to ageing and age-related pathologies. Presumably, some rare cells can escape from this normal default pathway by acquiring additional mutations that allow them to continue to proliferate even in the setting of damaged DNA. These proliferating but damaged cells might provide the seeds for future malignancies. In this scenario, both cancer and ageing result primarily from accumulating damage to the stem and progenitor cell compartment. Mutations that allow stem cells to continue to proliferate in the setting of normal growth arrest signals such as DNA damage (for example, loss of p16INK4a or reactivation of telomerase) would temporarily expand the stem cell pool and hence delay age-related pathologies. Over the long term, these mutations would also increase the likelihood of cancer.

undergo similar post-translational modification by molecules such as Sirt1 (Fig. 4).

Henrietta's gift

The complexity of ageing and the biology of cancer do not lend themselves to easy generalizations. As we have discussed, in some instances, such as cellular senescence or telomere shortening, strategies that protect us from cancer might hasten our rate of ageing. In other situations, such as autophagy or protection from genomic instability, cancer and ageing seem to share common, rather than antagonistic, aetiologies. In the coming years, it is to be hoped that researchers will further discern how and where these two entities coalesce and where their mechanisms diverge. Indeed, most of the fundamental questions remain unanswered. Among these is whether DNA damage represents the ultimate stimulus to both cancer and ageing. If so, is it the damage itself or the subsequent response to DNA damage that fuels ageing? Similarly, does the steep age-related increase in cancer incidence suggest that cancer is just one of a host of age-related pathologies? Or do the 'superactive' p53 mouse and the recent data on stem-cell deletion of the p16INK4a tumour suppressor indicate that mechanisms that limit the growth of cancer cells will also be rate-limiting for normal stem cells? If the latter is true, will any strategy to increase our molecular vigilance against cancer always come at the expense of limiting tissue regeneration and maintenance? Alternatively, can we view cancer and ageing as pure stem cell diseases, with cancer representing the effect of additional growth promoting mutations within a given stem cell and ageing representing the natural exhaustion and depletion of the stem and progenitor pool (Fig. 5)? Finally, will a deeper understanding of the molecular control of energy sensing and utilization, including what regulates mitochondrial activity and how the nucleus, mitochondria and cytoplasm communicate, provide new and fundamental insights into how we age and how a cancer cell emerges?

Undoubtedly, none of these questions were contemplated on that day in October 1951 when Henrietta Lacks's body was laid to rest in an unmarked grave near her family's small tobacco farm. Unbeknownst to those who gathered in that Virginia field—but as we now know—not all of Henrietta was buried that day. The small part that remained in the laboratory would forever change the course of science and help lead us to a clearer understanding of the barriers that separate normal cells from their cancer counterparts. These same barriers now appear to be intimately connected to how and why we age. Perhaps Henrietta's final gift to us is the growing realization that somewhere within the curse of the cancer cell's immortality there might also lie the secret of how we might understand and extend our own lifespan.

- Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp. Cell Res.* **25**, 585–621 (1961).
- Serrano, M., Hannon, G. J. & Beach, D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* **366**, 704–707 (1993).
- Alcorta, D. A. *et al.* Involvement of the cyclin-dependent kinase inhibitor p16 (INK4a) in replicative senescence of normal human fibroblasts. *Proc. Natl Acad. Sci. USA* **93**, 13742–13747 (1996).
- Hara, E. *et al.* Regulation of p16CDKN2 expression and its implications for cell immortalization and senescence. *Mol. Cell. Biol.* **16**, 859–867 (1996).
- Mooi, W. J. & Peeper, D. S. Oncogene-induced cell senescence—halting on the road to cancer. *N. Engl. J. Med.* **355**, 1037–1046 (2006).
- Collado, M. & Serrano, M. The power and the promise of oncogene-induced senescence markers. *Nature Rev. Cancer* **6**, 472–476 (2006).
- Bracken, A. P. *et al.* The Polycomb group proteins bind throughout the INK4AARF locus and are disassociated in senescent cells. *Genes Dev.* **21**, 525–530 (2007).
- Bartkova, J. *et al.* Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* **444**, 633–637 (2006).
- Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–642 (2006).
- Xue, W. *et al.* Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656–660 (2007).
- Ventura, A. *et al.* Restoration of p53 function leads to tumour regression *in vivo*. *Nature* **445**, 661–665 (2007).
- Roninson, I. B. Tumour cell senescence in cancer treatment. *Cancer Res.* **63**, 2705–2715 (2003).
- Gasser, S., Orsulic, S., Brown, E. J. & Raulet, D. H. The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature* **436**, 1186–1190 (2005).
- Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
- Herbig, U., Ferreira, M., Condel, L., Carey, D. & Sedivy, J. M. Cellular senescence in aging primates. *Science* **311**, 1257 (2006).
- Krishnamurthy, J. *et al.* Ink4a/Arf expression is a biomarker of aging. *J. Clin. Invest.* **114**, 1299–1307 (2004).
- Kim, W. Y. & Sharpless, N. E. The regulation of INK4/ARF in cancer and aging. *Cell* **127**, 265–275 (2006).
- Edwards, M. G. *et al.* Gene expression profiling of aging reveals activation of a p53-mediated transcriptional program. *BMC Genomics* **8**, 80 (2007).
- Melzer, D. *et al.* A common variant of the p16(INK4a) genetic region is associated with physical function in older people. *Mech. Ageing Dev.* published online 27 March 2007 (doi: 10.1016/j.mad.2007.03.005).
- McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
- Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
- Valk-Lingbeek, M. E., Bruggeman, S. W. & van Lohuizen, M. Stem cells and cancer: the polycomb connection. *Cell* **118**, 409–418 (2004).
- Janzen, V. *et al.* Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16INK4a. *Nature* **443**, 421–426 (2006).
- Krishnamurthy, J. *et al.* p16INK4a induces an age-dependent decline in islet regenerative potential. *Nature* **443**, 453–457 (2006).
- Molofsky, A. V. *et al.* Increasing p16INK4a expression decreases forebrain progenitors and neurogenesis during ageing. *Nature* **443**, 448–452 (2006).
- McMurray, M. A. & Gottschling, D. E. An age-induced switch to a hyperrecombinational state. *Science* **301**, 1908–1911 (2003).
- Curtis, H. J. Biological mechanisms underlying the aging process. *Science* **141**, 686–694 (1963).
- Bahar, R. *et al.* Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**, 1011–1014 (2006).
- Tyner, S. D. *et al.* p53 mutant mice that display early ageing-associated phenotypes. *Nature* **415**, 45–53 (2002).
- Maier, B. *et al.* Modulation of mammalian life span by the short isoform of p53. *Genes Dev.* **18**, 306–319 (2004).
- Garcia-Cao, I. *et al.* "Super p53" mice exhibit enhanced DNA damage response, are tumor resistant and age normally. *EMBO J.* **21**, 6225–6235 (2002).
- Matheu, A. *et al.* Delayed ageing through damage protection by the Arf/p53 pathway. *Nature* **448**, 375–379 (2007).
- Pinkston, J. M., Garigan, D., Hansen, M. & Kenyon, C. Mutations that increase the life span of *C. elegans* inhibit tumor growth. *Science* **313**, 971–975 (2006).
- Ito, K. *et al.* Regulation of oxidative stress by ATM is required for self-renewal of haematopoietic stem cells. *Nature* **431**, 997–1002 (2004).
- Cao, L., Li, W., Kim, S., Brodie, S. G. & Deng, C. X. Senescence, aging, and malignant transformation mediated by p53 in mice lacking the Brca1 full-length isoform. *Genes Dev.* **17**, 201–213 (2003).
- Baker, D. J. *et al.* BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. *Nature Genet.* **36**, 744–749 (2004).
- Baker, D. J. *et al.* Early aging-associated phenotypes in Bub3/Rae1 haploinsufficient mice. *J. Cell Biol.* **172**, 529–540 (2006).
- de Boer, J. *et al.* Premature aging in mice deficient in DNA repair and transcription. *Science* **296**, 1276–1279 (2002).
- Niedernhofer, L. J. *et al.* A new progeroid syndrome reveals that genotoxic stress suppresses the somatotrophic axis. *Nature* **444**, 1038–1043 (2006).
- Lombard, D. B. *et al.* DNA repair, genome stability, and aging. *Cell* **120**, 497–512 (2005).
- Opreko, P. L. *et al.* The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2. *Mol. Cell* **14**, 763–774 (2004).
- Agrelo, R. *et al.* Epigenetic inactivation of the premature aging Werner syndrome gene in human cancer. *Proc. Natl Acad. Sci. USA* **103**, 8822–8827 (2006).
- Liu, B. *et al.* Genomic instability in laminopathy-based premature aging. *Nature Med.* **11**, 780–785 (2005).
- Haigis, M. C. & Guarente, L. P. Mammalian sirtuins—emerging roles in physiology, aging, and calorie restriction. *Genes Dev.* **20**, 2913–2921 (2006).
- Vaquero, A. *et al.* Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol. Cell* **16**, 93–105 (2004).
- Pruitt, K. *et al.* Inhibition of SIRT1 reactivates silenced cancer genes without loss of promoter DNA hypermethylation. *PLoS Genet.* **2**, e40 (2006).
- Mostoslavsky, R. *et al.* Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* **124**, 315–329 (2006).
- Blackburn, E. H. & Gall, J. G. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J. Mol. Biol.* **120**, 33–53 (1978).
- Shampay, J., Szostak, J. W. & Blackburn, E. H. DNA sequences of telomeres maintained in yeast. *Nature* **310**, 154–157 (1984).
- Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413 (1985).

51. Lundblad, V. & Blackburn, E. H. An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell* **73**, 347–360 (1993).
52. Dunham, M. A., Neumann, A. A., Fasching, C. L. & Reddel, R. R. Telomere maintenance by recombination in human cells. *Nature Genet.* **26**, 447–450 (2000).
53. Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460 (1990).
54. Allsopp, R. C. *et al.* Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl Acad. Sci. USA* **89**, 10114–10118 (1992).
55. Cawthon, R. M. *et al.* Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* **361**, 393–395 (2003).
56. Bodnar, A. G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).
57. Blasco, M. A. *et al.* Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
58. Lee, H.-W., Blasco, M. A., Gottlieb, G. J., Greider, C. W. & DePinho, R. A. Essential role of mouse telomerase in highly proliferative organs. *Nature* **392**, 569–574 (1998).
59. Herrera, E. *et al.* Disease states associated to telomerase deficiency appear earlier in mice with short telomeres. *EMBO J.* **18**, 2950–2960 (1999).
60. Flores, I., Cayuela, M. L. & Blasco, M. A. Effects of telomerase and telomere length on epidermal stem cell behavior. *Science* **309**, 1253–1256 (2005).
61. Yamaguchi, H. *et al.* Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. *N. Engl. J. Med.* **352**, 1413–1424 (2005).
62. Armanios, M. Y. *et al.* Telomerase mutations in families with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* **356**, 1317–1326 (2007).
63. Tsakiri, K. D. *et al.* Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc. Natl Acad. Sci. USA* **104**, 7552–7557 (2007).
64. Mason, P. J., Wilson, D. B. & Bessler, M. Dyskeratosis congenita—a disease of dysfunctional telomere maintenance. *Curr. Mol. Med.* **5**, 159–170 (2005).
65. Garcia-Cao, I. *et al.* Increased p53 activity does not accelerate telomere-driven aging. *EMBO Rep.* **7**, 546–552 (2006).
66. Gonzalez-Suarez, E. *et al.* Increased epidermal tumors and increased skin wound healing in transgenic mice overexpressing the catalytic subunit of telomerase, mTERT, in basal keratinocytes. *EMBO J.* **20**, 2619–2630 (2001).
67. Artandi, S. E. *et al.* Constitutive telomerase expression promotes mammary carcinomas in aging mice. *Proc. Natl Acad. Sci. USA* **99**, 8191–8196 (2002).
68. Gonzalez-Suarez, E., Geserick, C., Flores, J. M. & Blasco, M. A. Antagonistic effects of telomerase on cancer and aging in K5-mTert transgenic mice. *Oncogene* **24**, 2256–2270 (2005).
69. Sarin, K. Y. *et al.* Conditional telomerase induction causes proliferation of hair follicle stem cells. *Nature* **436**, 1048–1052 (2005).
70. Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015 (1994).
71. Counter, C. M., Hirte, H. W., Bacchetti, S. & Harley, C. B. Telomerase activity in human ovarian carcinoma. *Proc. Natl Acad. Sci. USA* **91**, 2900–2904 (1994).
72. Bryan, T. M. *et al.* Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nature Med.* **3**, 1271–1274 (1997).
73. Blanco, R., Muñoz, P., Klatt, P., Flores, J. M. & Blasco, M. A. Telomerase abrogation dramatically accelerates TRF2-induced epithelial carcinogenesis. *Genes Dev.* **21**, 206–220 (2007).
74. Ohsumi, Y. Molecular dissection of autophagy: two ubiquitin-like systems. *Nature Rev. Mol. Cell Biol.* **2**, 211–216 (2001).
75. Liang, X. H. *et al.* Induction of autophagy and inhibition of tumorigenesis by beclin 1. *Nature* **402**, 672–676 (1999).
76. Yue, Z., Jin, S., Yang, C., Levine, A. J. & Heintz, N. Beclin 1, an autophagy gene essential for early embryonic development, is a haploinsufficient tumor suppressor. *Proc. Natl Acad. Sci. USA* **100**, 15077–15082 (2003).
77. Qu, X. *et al.* Promotion of tumorigenesis by heterozygous disruption of the beclin 1 autophagy gene. *J. Clin. Invest.* **112**, 1809–1820 (2003).
78. Kondo, Y. & Kondo, S. Autophagy and cancer therapy. *Autophagy* **2**, 85–90 (2006).
79. Kuma, A. *et al.* The role of autophagy during the early neonatal starvation period. *Nature* **432**, 1032–1036 (2004).
80. Komatsu, M. *et al.* Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature* **441**, 880–884 (2006).
81. Hara, T. *et al.* Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. *Nature* **441**, 885–889 (2006).
82. Melendez, A. *et al.* Autophagy genes are essential for dauer development and lifespan extension in *C. elegans*. *Science* **301**, 1387–1391 (2003).
83. Bensaad, K. *et al.* TIGAR, a p53-inducible regulator of glycolysis and apoptosis. *Cell* **126**, 107–120 (2006).
84. Matoba, S. *et al.* p53 regulates mitochondrial respiration. *Science* **312**, 1650–1653 (2006).
85. Sabatini, D. M. mTOR and cancer: insights into a complex relationship. *Nature Rev. Cancer* **6**, 729–734 (2006).
86. Wullschlegel, S., Loewith, R. & Hall, M. N. TOR signaling in growth and metabolism. *Cell* **124**, 471–484 (2006).
87. Velai, T. *et al.* Genetics: influence of TOR kinase on lifespan in *C. elegans*. *Nature* **426**, 620 (2003).
88. Kapahi, P. *et al.* Regulation of lifespan in *Drosophila* by modulation of genes in the TOR signaling pathway. *Curr. Biol.* **14**, 885–890 (2004).
89. Kaeberlein, M. *et al.* Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. *Science* **310**, 1193–1196 (2005).
90. Schieke, S. M. *et al.* The mammalian target of rapamycin (mTOR) pathway regulates mitochondrial oxygen consumption and oxidative capacity. *J. Biol. Chem.* **281**, 27643–27652 (2006).
91. Murphy, C. T. The search for DAF-16/FOXO transcriptional targets: approaches and discoveries. *Exp. Gerontol.* **41**, 910–921 (2006).
92. Hu, M. C. *et al.* IκB kinase promotes tumorigenesis through inhibition of forkhead FOXO3a. *Cell* **117**, 225–237 (2004).
93. Paik, J. H. *et al.* FoxOs are lineage-restricted redundant tumor suppressors and regulate endothelial cell homeostasis. *Cell* **128**, 309–323 (2007).
94. Tothova, Z. *et al.* FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128**, 325–339 (2007).
95. Baysal, B. E. *et al.* Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. *Science* **287**, 848–851 (2000).
96. King, A., Selak, M. A. & Gottlieb, E. Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer. *Oncogene* **25**, 4675–4682 (2006).
97. Ishii, N. *et al.* A mutation in succinate dehydrogenase cytochrome b causes oxidative stress and ageing in nematodes. *Nature* **394**, 694–697 (1998).
98. Walker, D. W. *et al.* Hypersensitivity to oxygen and shortened lifespan in a *Drosophila* mitochondrial complex II mutant. *Proc. Natl Acad. Sci. USA* **103**, 16382–16387 (2006).
99. Dillin, A. *et al.* Rates of behavior and aging specified by mitochondrial function during development. *Science* **298**, 2398–2401 (2002).
100. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33**, 40–48 (2003).
101. Levine, A. J., Feng, Z., Mak, T. W., You, H. & Jin, S. Coordination and communication between the p53 and IGF-1-AKT-TOR signal transduction pathways. *Genes Dev.* **20**, 267–275 (2006).

Acknowledgements We thank members of our laboratory for helpful discussions and I. Rovira for help with the manuscript. This work was supported by grants from the NIH Intramural program and Ellison Medical Foundation (T.F.), the Spanish Ministry of Education and Science (M.S., M.A.B.), the European Union (M.S., M.A.B.) and the Josef Steiner Award (M.A.B.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to T.F. (finkelt@nih.gov).

Structure-based activity prediction for an enzyme of unknown function

Johannes C. Hermann¹, Ricardo Marti-Arbona², Alexander A. Fedorov³, Elena Fedorov³, Steven C. Almo³, Brian K. Shoichet¹ & Frank M. Raushel²

With many genomes sequenced, a pressing challenge in biology is predicting the function of the proteins that the genes encode. When proteins are unrelated to others of known activity, bioinformatics inference for function becomes problematic. It would thus be useful to interrogate protein structures for function directly. Here, we predict the function of an enzyme of unknown activity, Tm0936 from *Thermotoga maritima*, by docking high-energy intermediate forms of thousands of candidate metabolites. The docking hit list was dominated by adenine analogues, which appeared to undergo C6-deamination. Four of these, including 5-methylthioadenosine and S-adenosylhomocysteine (SAH), were tested as substrates, and three had substantial catalytic rate constants ($10^5 \text{ M}^{-1} \text{ s}^{-1}$). The X-ray crystal structure of the complex between Tm0936 and the product resulting from the deamination of SAH, S-inosylhomocysteine, was determined, and it corresponded closely to the predicted structure. The deaminated products can be further metabolized by *T. maritima* in a previously uncharacterized SAH degradation pathway. Structure-based docking with high-energy forms of potential substrates may be a useful tool to annotate enzymes for function.

For enzymes of unknown function, substrate prediction based on structural complementarity is, in principle, an alternative to bioinformatics inference of function^{1,2}. Structure-based prediction becomes attractive when the target enzyme has little relationship to orthologues of known activity, making inference unreliable^{3,4}. Whereas structure-based prediction has been used with some successes for inhibitor design, substrate prediction has proven difficult^{5–8}. In addition to the well-known problems of sampling and scoring in docking, substrate prediction confronts several additional challenges. These include the many possible substrates to consider and the many reactions that an enzyme might catalyse^{9–11}. Furthermore, enzymes preferably recognize transition states over the ground state structures that are usually represented in docking databases^{12–14}.

Docking metabolites as high-energy intermediates

If, in its most general form, structure-based substrate prediction seems daunting, it may be simplified by several pragmatic choices. If we focus on a single class of reactions, here those catalysed by the amidohydrolase superfamily (AHS), of which Tm0936 is a member, we reduce the number of possible reactions from practically unbounded to a limited set of mechanistically related transformations. Thus, the 6,000 catalogued members of the AHS catalyse ~30 reactions in biosynthetic and catabolic pathways^{15–17}. All adopt a common (β/α)₈-barrel fold and almost all are metallo-enzymes that cleave carbon-hetero-atom bonds. The problem of activity prediction may be further simplified by focusing on a single source of likely substrates, here the KEGG metabolite database¹⁸. Although substrate identification remains challenging—there are probably hundreds of molecules that are specifically recognized, not all of which are metabolites—it is at least a finite problem.

To address the challenge of transition state recognition, ground state structures were transformed into structures mimicking the

high-energy intermediates that occur along the enzyme reaction coordinate. We will refer to these transition-state-like geometries as high-energy intermediates; this form of the substrate is among those that should best complement steric and electronic features of the enzyme active site^{14,19}. All functional groups potentially recognized by AHS enzymes, for each of the 4,207 metabolites that bore them, were converted into high-energy intermediate geometries, with their appropriate charge distributions (Fig. 1). For instance, aromatic amines, which in the ground state are planar, are converted computationally into tetrahedral centres, representing the high-energy intermediate for deamination. Similarly, tetrahedral phosphates are converted into trigonal, bipyramidal forms. Overall, 28 amidohydrolase reactions operating on 19 functional groups were modelled by these high-energy structures, leading to the calculation of about 22,500 different forms of the metabolites. In retrospective calculations, docking these high-energy intermediate structures into seven well-studied amidohydrolases consistently identified the correct substrate from among the thousands of decoy molecules, typically outperforming docking of the ground state forms of the same molecules^{20,21}.

These retrospective results encouraged us to prospectively predict the substrates of Tm0936 from *T. maritima*. The X-ray structure of the enzyme had been determined as part of a broad structural genomics effort (PDB codes 1p1m and 1j6p), and it can be assigned to the AHS by fold classification and the identity of certain active site groups. Despite this, its substrate preference is anything but clear. By sequence similarity, Tm0936 most resembles the large chlorohydrolase and cytosine deaminase subgroup, which is often used to annotate amidohydrolases of unknown function¹⁷. Consistent with the view that this reflects an assignment to a broad subfamily and not a functional annotation, we tested 14 cytosine derivatives as Tm0936 substrates; no turnover was detected for any of them (see Methods). In an effort to find the true substrate, we therefore docked the

¹Department of Pharmaceutical Chemistry, University of California, San Francisco, MC 2550 1700 4th Street, San Francisco, California 94158-2330, USA. ²Department of Chemistry, P.O. Box 30012, Texas A&M University, College Station, Texas 77842-3012, USA. ³Department of Biochemistry, Albert Einstein College of Medicine, Ullmann Building, Room 411, 1300 Morris Park Avenue, Bronx, New York 10461, USA.

database of high-energy intermediates into the structure of Tm0936, sampling thousands of configurations and conformations of each molecule. Each of these was scored by electrostatic and van der Waals complementarity, corrected for ligand desolvation energy, and ranked accordingly (see Methods)^{22,23}.

The molecules best-ranked computationally were dominated by adenine and adenosine analogues, which make up 9 of the 10 top-scoring docking hits (Table 1, Supplementary Fig. 1). For all of these, an exocyclic nitrogen has been transformed into a tetrahedral, high energy centre, as would occur in a deamination reaction. The dominance of adenine and adenosine analogues, in this form, is due to nearly ideal interactions with the active site. An example is the docked structure of the high-energy intermediate for the deamination of 5-methylthioadenosine (MTA), the 6th ranked molecule (Fig. 2).

Experimental testing of the predicted substrates

On the basis of the docking ranks and compound availability, we selected four potential substrates for deamination by Tm0936: MTA, SAH, adenosine and adenosine monophosphate (AMP), all of which scored well (5th, 6th, 14th, 80th out of 4,207 docked metabolites), underwent the same reaction, and chemically resembled one another (Table 2). Although there were other high-ranking molecules in the docking hit list, most were single representatives of a chemotype and lacked the virtue of consistency of the adenines in general and the adenosines in particular. By extension, we also investigated the well-known metabolite *S*-adenosyl-L-methionine (SAM), a close analogue of SAH, even though its docking rank, at 511th, was poor.

Of these five molecules, three had substantial activity as substrates, with MTA and SAH reaching k_{cat}/K_m values of 1.4×10^5 and $5.8 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ respectively, and adenosine close to $10^4 \text{ M}^{-1}\text{s}^{-1}$ (Table 2 and Supplementary Information). The first order rate constant for the spontaneous deamination of adenosine in water is $1.8 \times 10^{-10} \text{ s}^{-1}$, making this enzyme proficient for these substrates. Tm0936 is relatively active compared to other adenosine deaminases²⁴, especially because the optimal temperature for this thermophilic enzyme is almost certainly higher than the 30 °C at which it was assayed. Consistent with the docking predictions, SAM was not deaminated by Tm0936, despite its close similarity to SAH. Conversely, AMP, which did rank relatively well (80th of 4,207), was also not an enzyme substrate. The inability of the docking programme to fully de-prioritize AMP reflects some of the well-known problems in docking scoring functions, in this case balancing ionic interactions and desolvation penalties for the highly charged phosphate group of AMP.

To investigate the mechanism further, we determined the structure of Tm0936 in complex with the purified product of the SAH deamination reaction, *S*-inosylhomocysteine (SIH), to 2.1 Å resolution by X-ray crystallography (Fig. 3, Methods). The differences between the docked prediction and the crystallographic result are minor, with every key polar and non-polar interaction represented in both structures (except that we docked the tetrahedral intermediate and the X-ray structure is of the ground state product). Indeed, the

correspondence between the docked and crystallographic structures is closer than one might expect for inhibitor predictions, where docking has been more commonly used^{25–28}. This may reflect the advantages of docking substrates in high-energy intermediate geometries, which encode more of the information necessary to specify fit.

Metabolic pathway of a family of MTA/SAH deaminases

It is tempting to speculate that Tm0936 is not simply an isolated enzyme acting on particular substrates, but is involved in the deamination of metabolites in a previously uncharacterized MTA/SAH pathway. The deamination of adenosine itself is well known in all kingdoms of life, and the deamination of SAH to SIH has been reported in one organism, *Streptomyces flocculus*²⁹. Very recently it was shown that MTA is deaminated in *Plasmodium falciparum* in an alternative degradation pathway of adenosine analogues³⁰. To investigate whether the products of the deamination reactions, catalysed by Tm0936, SIH and MTI, could be further metabolized by other

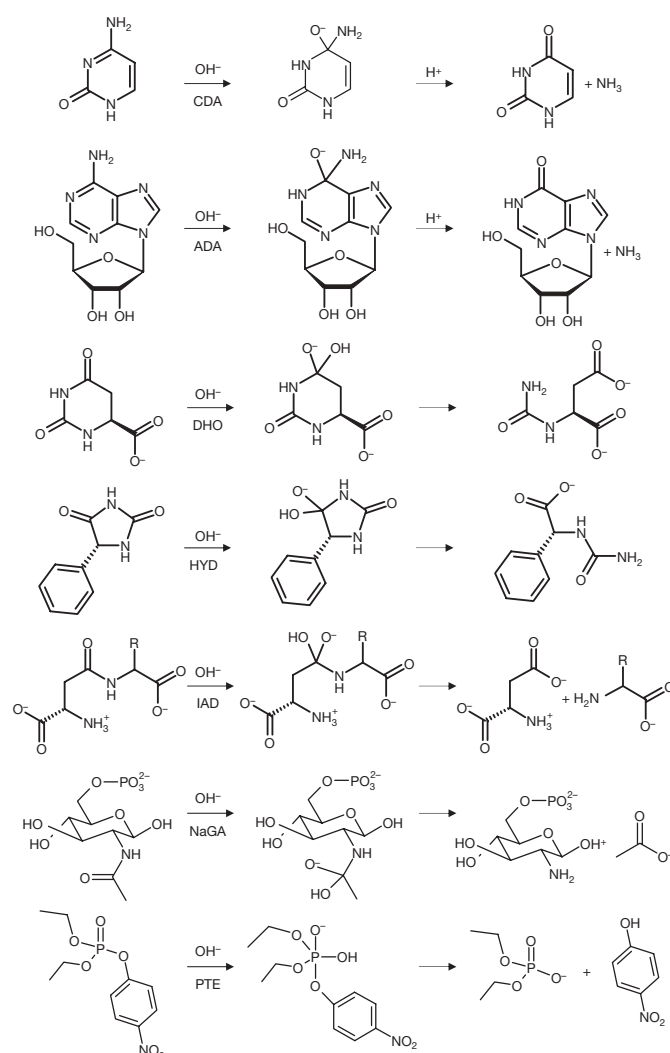


Figure 1 | Sample transformations of metabolites from their ground state structure into the high-energy intermediate forms that were used for docking. Transformations were computed according to the conserved reaction mechanism of amidohydrolases, a nucleophilic attack of a hydroxide at an electrophilic centre atom. Every transformable functional group for each molecule was processed independently. If the high-energy structure was chiral, all stereoisomers were calculated. Reactions catalysed by the amidohydrolases cytosine deaminase (CDA), adenosine deaminase (ADA), dihydroorotase (DHO), D-hydantoinase (HYD), isoaspartyl-D-dipeptidase (IAD), *N*-acetyl-D-glucosamine-6-phosphate deacetylase (NaGA) and phosphotriesterase (PTE) are shown.

Table 1 | The occurrence of adenine analogues among the top-ranked docking hits

Analogues in docking hit list	Top 10 ranked hits	Top 20 ranked hits	Top 100 ranked hits	Top 300 ranked hits
Adenine analogues				
	9	17	32	44
Enrichment factor	34	32	12	6

The enrichment factor is measured relative to the abundance of the analogues among the 4,207 potential substrates docked.

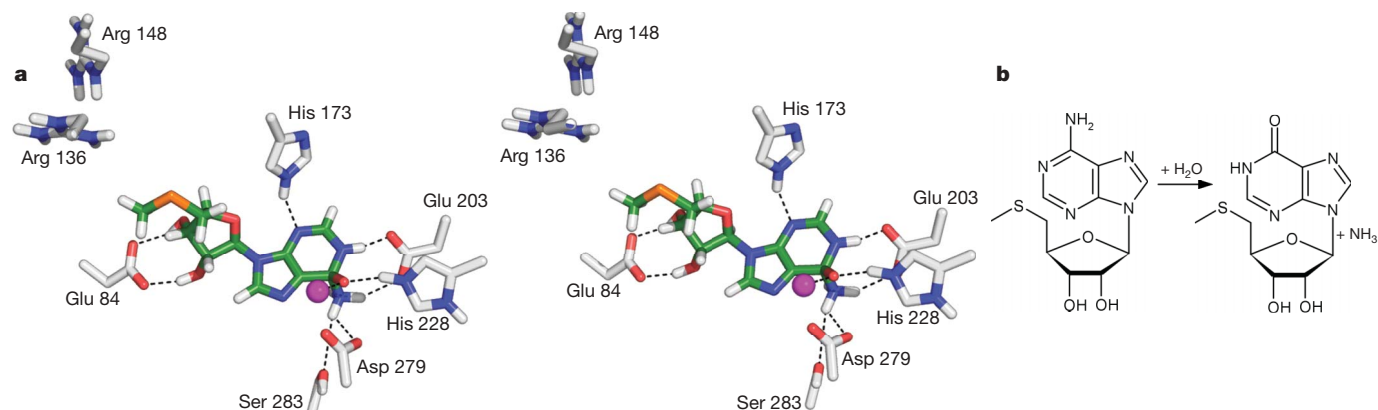


Figure 2 | Binding and conversion of MTA by Tm0936. a, Stereoview of MTA in its high-energy intermediate form docked into the active site of Tm0936. Oxygen atoms are coloured red; enzyme carbons, grey; ligand carbons, green; hydrogens, white; nitrogens, blue; sulphur, orange; and the metal ion, purple. The oxyanion, representing the nucleophilic hydroxyl, ion-pairs with the metal ion and His 228. The ammonia leaving group is placed between Glu 203 and Asp 279, at 3.2 Å and 2.9 Å, respectively, also

interacting with Ser 283 (3.2 Å). The N1-nitrogen donates a hydrogen bond to Glu 203, whereas N3 accepts one from His 173 (2.5 Å and 2.9 Å). Ribose hydroxyls hydrogen bond to Glu 84 (2.8 Å and 2.9 Å). Adenosines larger than MTA, such as SAH, make additional interactions with more distal residues, such as Arg 136 and Arg 148. All figures were rendered using PyMOL (<http://pymol.sourceforge.net>). **b,** The deamination of MTA to MTI, a reaction catalysed by Tm0936.

Table 2 | Docking ranks and Tm0936 catalytic constants for five predicted substrates

Substrate tested	Docked high-energy intermediate form	Dock rank	Relative docking scores (kcal mol ⁻¹)*	K_m (μM)	k_{cat} (s ⁻¹)	k_{cat}/K_m (M ⁻¹ s ⁻¹)
S-adenosyl-L-homocysteine		5	0	210 ± 40	12.2 ± 0.8	5.8 × 10 ⁴
5-Methyl-thioadenosine		6	4.4	44 ± 4	7.2 ± 0.2	1.4 × 10 ⁵
Adenosine		14	9.5	250 ± 40	2.3 ± 0.2	9.2 × 10 ³
Adenosine-5-monophosphate		80	20.2	ND	<10 ⁻³	ND
S-adenosyl-L-methionine		511	35.2	ND	<10 ⁻³	ND

Deamination was measured by the production of ammonia. The standard deviations are given.

* Docking energies relative to the best-ranked compound shown, SAH. Higher energies indicate worse scores. ND, not determined.

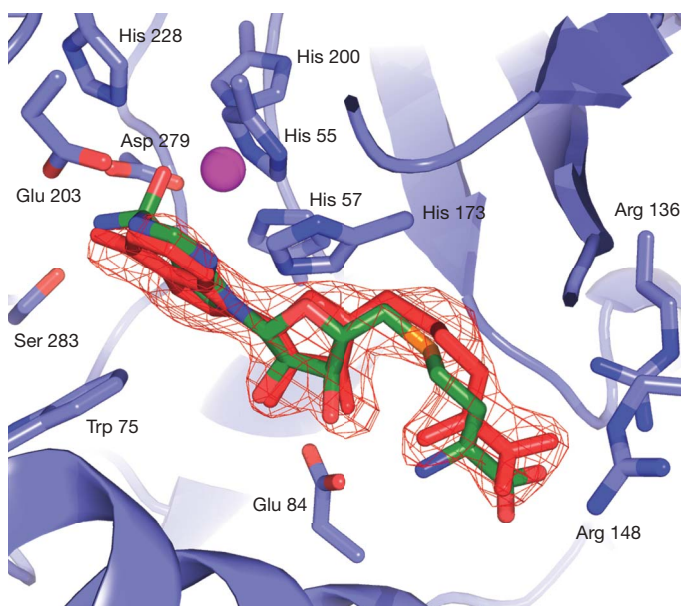


Figure 3 | Comparing the docking prediction and the crystallographic result. Superposition of the crystal structure of Tm0936 in complex with SIH (red) and the docking predicted structure of the high-energy intermediate of SAH (carbons in green). Enzyme carbons are coloured light blue, SAH and enzyme oxygen atoms are coloured red, nitrogens blue and sulphurs orange. The purple sphere represents the divalent metal ion. An $F_o - F_c$ omit electron density map for SIH is shown, contoured at 4.1 σ . The structure was determined at 2.1 Å resolution.

enzymes in *T. maritima*, we measured the activity of S-adenosyl homocysteinase (Tm0172), which hydrolyses SAH to homocysteine and adenosine, using SIH as a potential substrate. We found that Tm0172 catalyses the formation of homocysteine from either SIH or SAH about equally well (Supplementary Table 1 and Supplementary Information). This is consistent with Tm0172 and Tm0936 participating in a degradation pathway, though it does not confirm it. We cannot exclude the possibility that Tm0936 functions as an adenosine deaminase in *T. maritima*, because no other enzyme in the organism has been identified that serves this role.

What is clear is that Tm0936 has orthologues across multiple species. On the basis of the conservation of characteristic residues that interact with the substrate and product in the docked and X-ray structures, respectively, 78 other previously unannotated AHS enzymes from different species may now be classified as MTA/SAH/adenosine deaminases (Supplementary Fig. 2 and Supplementary Information). In all of these sequences, the metal-ligating residues (His 55, His 57, His 200 and Asp 279, Tm0936 numbering) are conserved, as are the residues recognizing the reactive centre (His 228, Ser 259, Ser 283 and Glu 203). Specificity is conferred by interactions between the substrate and Trp 75, Glu 84 and His 173, all of which are also conserved among the 78 amidohydrolases. Active site residues that vary include Arg 136 and Arg 148, which in Tm0936 interact with the α -carboxylate of the homocysteine moiety of SAH. These latter interactions are not critical to the activity of the enzyme, because these arginines do not seem to interact with MTA or adenosine, but they may be important for the recognition of SAH.

Many of the Tm0936 orthologues cluster with other genes that can now be associated with the metabolism of SAM, SAH or MTA. For example, in *T. maritima* Tm0936 is closely associated with Tm0938, which is currently annotated as a SAM-dependent methyl transferase. In *Bacillus cereus*, the Tm0936 orthologue is Bc1793, which is also closely associated with a SAM-dependent methyl transferase, Bc1797. In *Pseudomonas aeruginosa*, the Tm0936 orthologue, Pa3170, is adjacent to UbiG-methyltransferase, Pa3171. Other orthologues are adjacent or close to adenosyl homocysteinase, 5'-methylthioadenosine

phosphorylase, MTA/SAH nucleosidase and other SAM-dependent methyl transferases.

Predicting function from form

This work describes one case of successful function prediction by structure-based docking, and it is appropriate to consider caveats. Our recognition of Tm0936 as an amidohydrolase limited the number of possible reactions to be considered. When even the gross mechanistic details of an enzyme cannot be inferred, this will not be possible. Restricting ourselves to metabolites was also helpful, but this too will not always be appropriate. Finally, we were fortunate that Tm0936 experienced little conformational change between the apo structure and that of the product complex. Enzymes that undergo large conformational changes along their reaction coordinates will be more challenging for docking.

If prudence warns against over-generalization, it is also unlikely that Tm0936 represents an isolated case. Other enzyme structures will be broadly classifiable by mechanism, and whereas conformational change remains a serious challenge, retrospective studies suggest that it is not insurmountable. Indeed, the most important technical innovation adopted here, modelling substrates as high-energy intermediates, was particularly useful when docking to apo structures in those studies (Supplementary Table 2 and Supplementary Information)²⁰. Thus, the prediction and determination that Tm0936 acts as an MTA/SAH deaminase illustrates the possibilities of this and related structure-based approaches, at least for a subset of targets. The enzyme has no obvious sequence similarity to any known adenosine deaminase and exploits interactions not previously identified in the active sites of these enzymes. The very pathway in which Tm0936 participates seems novel. Structure-based docking of high-energy intermediates should be a useful tool to decrypt the activity of enzymes of unknown function, and will be especially interesting for those targets where bioinformatics inference is unreliable.

METHODS

Molecular docking. The 1.5 Å X-ray structure of Tm0936 (Protein Data Bank (PDB) code 1P1M) was used in docking calculations. High-energy intermediates of potential substrates were calculated²⁰ and docked into the enzyme structure using the program DOCK3.5.54. Poses were scored for electrostatic and van der Waals complementarity and penalized for ligand desolvation^{31,32}.

Enzymology. Tm0936 and Tm0172 from *T. maritima* were cloned, expressed and purified using standard techniques. The deamination reaction was measured by coupling the production of ammonia to the oxidation of NADH catalysed by glutamate dehydrogenase. The decrease in the concentration of NADH was followed spectrophotometrically at 340 nm. The chemical identities of the deaminated products were confirmed by mass spectrometry and specific changes in the ultraviolet absorption (UV) spectra for the deamination of adenosine derivatives. The SAH hydrolase activity by Tm0172 was determined by reaction of the free thiol group of the homocysteine product with dithio-bis(2-nitrobenzoic acid), monitoring the absorbance at 412 nm.

X-ray crystallography. Tm0936 was co-crystallized with ZnCl₂ and SIH. X-ray diffraction data were collected at the NSLS X4A beamline (Brookhaven National Laboratory). The structure of the Tm0936-SIH complex was determined by molecular replacement, using apo Tm0936 (PDB code 1J6P) as the search model. The structure has been deposited in the protein data bank (PDB code 2PLM).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 January; accepted 7 June 2007.

Published online 1 July 2007.

- Whisstock, J. C. & Lesk, A. M. Prediction of protein function from protein sequence and structure. *Q. Rev. Biophys.* **36**, 307–340 (2003).
- Gerlt, J. A. & Babbitt, P. C. Can sequence determine function? *Genome. Biol.* **1**, REVIEWS0005 (2000).
- Brenner, S. E. Errors in genome annotation. *Trends Genet.* **15**, 132–133 (1999).
- Devos, D. & Valencia, A. Intrinsic errors in genome annotation. *Trends Genet.* **17**, 429–431 (2001).
- Schapira, M., Abagyan, R. & Totrov, M. Nuclear hormone receptor targeted virtual screening. *J. Med. Chem.* **46**, 3045–3059 (2003).

6. Rao, M. S. & Olson, A. J. Modelling of factor Xa-inhibitor complexes: a computational flexible docking approach. *Proteins* **34**, 173–183 (1999).
7. Sukuru, S. C. *et al.* Discovering new classes of *Brugia malayi* asparaginyl-tRNA synthetase inhibitors and relating specificity to conformational change. *J. Comput. Aided. Mol. Des.* **20**, 159–178 (2006).
8. Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432**, 862–865 (2004).
9. Macchiarulo, A., Nobeli, I. & Thornton, J. M. Ligand selectivity and competition between enzymes *in silico*. *Nature Biotechnol.* **22**, 1039–1045 (2004).
10. Kalyanaraman, C., Bernacki, K. & Jacobson, M. P. Virtual screening against highly charged active Sites: identifying substrates of α - β barrel enzymes. *Biochemistry* **44**, 2059–2071 (2005).
11. Irwin, J. J. & Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177–182 (2005).
12. Schramm, V. L. Enzymatic transition states and transition state analogues. *Curr. Opin. Struct. Biol.* **15**, 604–613 (2005).
13. Hermann, J. C., Ridder, L., Holtje, H. D. & Mulholland, A. J. Molecular mechanisms of antibiotic resistance: QM/MM modelling of deacylation in a class A β -lactamase. *Org. Biomol. Chem.* **4**, 206–210 (2006).
14. Warshel, A. & Florian, J. Computer simulations of enzyme catalysis: finding out what has been optimized by evolution. *Proc. Natl Acad. Sci. USA* **95**, 5950–5955 (1998).
15. Holm, L. & Sander, C. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins* **28**, 72–82 (1997).
16. Seibert, C. M. & Raushel, F. M. Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* **44**, 6383–6391 (2005).
17. Pegg, S. C. *et al.* Leveraging enzyme structure–function relationships for functional inference and experimental design: the structure–function linkage database. *Biochemistry* **45**, 2545–2555 (2006).
18. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
19. Tantillo, D. J. & Houk, K. N. Transition state docking: a probe for noncovalent catalysis in biological systems. Application to antibody-catalyzed ester hydrolysis. *J. Comput. Chem.* **23**, 84–95 (2002).
20. Hermann, J. C. *et al.* Predicting substrates by docking high-energy intermediates to enzyme structures. *J. Am. Chem. Soc.* **128**, 15882–15891 (2006).
21. Nowlan, C. *et al.* Resolution of chiral phosphate, phosphonate, and phosphinate esters by an enantioselective enzyme library. *J. Am. Chem. Soc.* **128**, 15892–15902 (2006).
22. Wei, B. Q., Baase, W. A., Weaver, L. H., Matthews, B. W. & Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* **322**, 339–355 (2002).
23. Lorber, D. M. & Shoichet, B. K. Hierarchical docking of databases of multiple ligand conformations. *Curr. Top. Med. Chem.* **5**, 739–749 (2005).
24. Radzicka, A. & Wolfenden, R. A proficient enzyme. *Science* **267**, 90–93 (1995).
25. Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P. & DesJarlais, R. L. Docking: successes and challenges. *Curr. Pharm. Des.* **11**, 323–333 (2005).
26. Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **303**, 1813–1818 (2004).
27. Kairys, V., Fernandes, M. X. & Gilson, M. K. Screening drug-like compounds by docking to homology models: a systematic study. *J. Chem. Inf. Model.* **46**, 365–379 (2006).
28. Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today* **11**, 580–594 (2006).
29. Speedie, M. K., Zulty, J. J. & Brothers, P. S-adenosylhomocysteine metabolism in *Streptomyces flocculus*. *J. Bacteriol.* **170**, 4376–4378 (1988).
30. Tyler, P. C., Taylor, E. A., Fröhlich, R. F. G. & Schramm, V. L. Synthesis of 5'-methylthio coformycins: specific inhibitors for malarial adenosine deaminase. *J. Am. Chem. Soc.* **129**, 6872–6879 (2007).
31. Meng, E. C., Shoichet, B. & Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comp. Chem.* **13**, 505–524 (1992).
32. Gschwend, D. A. & Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. *J. Comput. Aided Mol. Des.* **10**, 123–132 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institutes of Health, supporting docking analyses (to B.K.S.), large scale structural analysis (to S.C.A.), and function prediction (to F.M.R., B.K.S. and S.C.A.). F.M.R. thanks the Robert A. Welch Foundation for support. J.C.H. thanks the Deutsche Akademie der Naturforscher Leopoldina for a fellowship. We thank J. Irwin, V. Thomas and K. Babaoğlu for reading this manuscript. The clone for Tm0172 was kindly supplied by the Joint Center for Structural Genomics.

Author Contributions J.C.H. designed the docking database, performed the docking runs, and analysed the docking results. F.M.R. and R.M.-A. performed the enzymatic characterization of Tm0936 and Tm0172, including cloning and purification of the proteins. S.C.A., E.F. and A.A.F. determined the X-ray structure of Tm0936 with S-inosyl-homocysteine. J.C.H. and B.K.S. largely wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information The complex structure of Tm0936 with SIH has been deposited in the PDB (accession code 2PLM). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials related to docking should be addressed to B.K.S. (shoichet@cgl.ucsf.edu).

METHODS

Molecular docking. The 1.5 Å X-ray structure of Tm0936 was used in the docking calculations (PDB code 1P1M). The active site metal ion was assigned a charge of +1.4, the remaining charge of 0.6 was distributed among the ligating residues, His 55, His 57, His 200 and Asp 279, to keep the correct net charge and to account for charge distribution effects in metal complexes^{20,33}. His 228 was protonated according to its assumed function as the base to activate the catalytic water molecule by abstracting a proton; the hydroxide ion itself was removed from the active site, because it is part of each high-energy intermediate structure that we dock²⁰.

The programs CHEMGRID and DISTMAP were used to compute docking grids for van der Waals potentials and excluded volume, respectively³¹. The electrostatic potential grid was calculated with DELPHI using an internal dielectric of 2 and an external dielectric of 78 (ref. 34). A manually curated set of spheres based on a set calculated by the program SPHGEN was used to orient molecules in the binding site³⁶.

High-energy intermediates of potential substrates were docked into Tm0936 using the docking program DOCK3.5.54 (ref. 23). Initial ligand orientations were sampled using receptor and ligand bin sizes of 0.5 Å and a ligand and receptor overlap of 0.4 Å. A distance tolerance of 1.5 Å was used for matching receptor and ligand spheres. An average of more than a million poses per molecule were calculated, and those that sterically fit the site were scored for electrostatic and van der Waals complementarity and penalized for ligand desolvation³². The best scoring orientation was rigid-body-minimized according to these energies. The details of the preparation of high-energy structures for docking, the molecular docking procedure, and methods for the analysis of the results have been previously described²⁰.

Enzymatic characterization of Tm0936. All compounds and coupling enzymes were obtained from Sigma or Aldrich, unless otherwise specified. The genomic DNA from *T. maritima* was purchased from the American Type Culture Collection (ATCC). The oligonucleotide synthesis and DNA sequencing reactions were performed by the Gene Technology Laboratory of Texas A&M University. The pET30a(+) expression vector was acquired from Novagen. The T4 DNA ligase and the restriction enzymes, *NdeI* and *EcoRI*, were purchased from New England Biolabs. The Platinum *Pfx* DNA polymerase and the Wizard Plus SV Mini-Prep DNA purification kit were obtained from Invitrogen and Promega, respectively. The glycerol stock of the plasmid encoding Tm0172 was kindly provided by the Joint Center for Structural Genomics.

Cloning of Tm0936. The gene encoding Tm0936 from *Thermotoga maritima* was amplified from the genomic DNA by standard PCR methods stipulated in the manufacturer's instructions using oligonucleotide primers with *NdeI* and *EcoRI* restriction sites at either end (Supplementary Table 3). The PCR products were purified, digested with *NdeI* and *EcoRI*, ligated to the expression vector pET30a(+) using T4 DNA ligase, and then transformed into XL1Blue cells. Individual colonies containing the plasmid were selected on LB plates containing 50 µg ml⁻¹ kanamycin and then used to inoculate 5 ml cultures of LB. The entire coding regions of the plasmids containing the *Tm0936* gene were sequenced to confirm the fidelity of the PCR amplification.

Purification of Tm0936. Cells harbouring the plasmid for the expression of Tm0936 were grown overnight and a single colony was used to inoculate 50 ml of LB media containing 50 µM kanamycin, and subsequently used to inoculate 2 l of the same medium. Cell cultures were grown at 37 °C with a rotary shaker until an A_{600} of ~0.6 was reached. Induction was initiated by the addition of 1.0 mM isopropyl-thiogalactoside (IPTG), and further incubated overnight at 30 °C. The bacterial cells were isolated by centrifugation at 5,200 × *g* for 15 min at 4 °C. The pellet was re-suspended in 50 mM HEPES buffer, pH 7.5 (buffer A), containing 5 µg ml⁻¹ RNase and 0.1 mg ml⁻¹ PMSF per gram of wet cells and then disrupted by sonication. The soluble protein was separated from the cell debris by centrifugation at 14,000 × *g* for 15 min and heated at 65 °C for 15 min to precipitate the *Escherichia coli* proteins. The soluble protein was separated from the precipitated protein by centrifugation at 14,000 × *g* for 15 min, loaded onto a 6 ml Resource Q anion ion exchange column (GE Health Care) and eluted with a gradient of NaCl in 20 mM HEPES, pH 8.5 (buffer B). The fractions containing Tm0936 were pooled and re-precipitated by saturation with ammonium sulphate, centrifuged at 14,000 × *g* for 15 min at 4 °C, and resuspended in a minimum amount of buffer A. The final step in the purification was accomplished by chromatography on a High Load 26/60 Superdex 200 prep grade gel filtration column (GE Health Care) and eluted with buffer A. The purity of the protein during the isolation procedure was monitored by SDS-PAGE.

Purification of Tm0172. Cells harbouring the plasmid for the expression of Tm0172 were grown overnight and a single colony was used to inoculate 50 ml of LB media containing 100 µM ampicillin and subsequently used to inoculate 2 l of the same medium. Cell cultures were grown at 37 °C with a rotary

shaker until an A_{600} of ~0.6 was reached. Induction was initiated by the addition of 1.0 mM arabinose, and further incubated overnight at 37 °C. The bacterial cells were isolated by centrifugation at 5,200 × *g* for 15 min at 4 °C. The pellet was re-suspended in 20 mM Tris-Cl buffer, 5 mM imidazole and 500 mM NaCl at pH 7.5 (buffer A), containing 0.1 mg ml⁻¹ phenylmethylsulphonyl fluoride per gram of wet cells and then disrupted by sonication. The soluble protein was separated from the cell debris by centrifugation at 14,000 × *g* for 15 min and heated at 65 °C for 15 min to precipitate the *E. coli* proteins. The soluble protein was separated from the precipitated protein by centrifugation at 14,000 × *g* for 15 min, loaded onto a Chelating Sepharose Fast Flow column for histidine-tagged fusion protein purification and eluted with a gradient of imidazole in buffer A. Fractions containing the desired protein were pooled by catalytic activity and purity. The purity of the protein during the isolation procedure was monitored by SDS-PAGE.

Metal analysis and amino acid sequence verification. The purified Tm0936 was subjected to amino-terminal amino acid sequence analysis by the Protein Chemistry Laboratory at Texas A&M University. The first five amino acids were MIIGN, which agrees with the protein sequence reported for Tm0936. The protein concentration was determined spectrophotometrically at 280 nm using a SPECTRAMax-340 microplate reader (Molecular Devices). An extinction coefficient of 51,020 M⁻¹cm⁻¹ was used for Tm0936 on the basis of the protein sequence. The metal content of the purified protein was determined by inductively coupled plasma emission-mass spectrometry (ICP-MS) and found to contain 1.2 equivalents of Zn per subunit.

Determination of SAH deaminase activity. The measurement of the deaminating properties of Tm0936 was conducted by coupling the production of ammonia to the oxidation of NADH with glutamate dehydrogenase. The decrease in the concentration of NADH was followed spectrophotometrically at 340 nm using a SPECTRAMax-340 microplate reader. The standard assay was modified from the report in ref. 36, and contained 100 mM HEPES at pH 8.0, 7.4 mM α-ketoglutarate, 0.4 mM NADH, 6 units of glutamate dehydrogenase, Tm0936 and the appropriate compound in a final volume of 250 µl at 30 °C. Following the initial, purely bioinformatic predictions of cytosine deaminase activity, the following compounds were tested for enzymatic activity at a concentration of 10 mM using this protocol: cytosine, 5-methylcytosine, 5-fluorocytosine, 6-aminouracil, 4,6-diamino-2-hydroxypyrimidine, 2-deoxycytidine, cytosine-β-D-arabinofuranoside, cytidine, cytidine-5'-diphosphocholine, cytidine-5'-monophosphate, 2'-deoxycytidine-5'-diphosphate, cytidine-5'-diphosphate, cytidine-5'-triphosphate, cytidine-3'-phosphate. Subsequently, we cast a wider net looking for activity on N-formimino-L-glutamate, N-formimino-L-aspartate, N-formimino-L-glycine. It was only with the structure-based docking predictions that we turned to deamination of adenosines, first directly testing the docking predicted metabolites adenosine, adenosine-5'-monophosphate, 5'-methylthioadenosine, S-adenosine-5'-homocysteine. Eventually we tested also several other analogues including adenosine-5'-diphosphate, adenosine-5'-triphosphate, S-adenosine-5'-methionine, folate, thiamine, pterin, and guanine. Only adenosine, S-adenosine-5'-homocysteine, and 5'-methylthioadenosine were found to be substrates.

The products of the reaction catalysed by Tm0936 were confirmed by mass spectroscopy and by changes in the UV spectra. When S-adenosine-5'-homocysteine was incubated with Tm0936, the mass spectral signal for SAH at a [M+H]⁺ of 385 *m/z* disappeared and was replaced by a new signal at a [M+H]⁺ of 386 *m/z* that corresponds to the mass expected for S-inosyl-5'-homocysteine (SIH). The UV spectrum for SAH has a maximal absorbance at 260 nm and after the addition of Tm0936 the absorbance maximum shifts to 250 nm. These results are consistent with the deamination of the adenine moiety of the substrate and conversion to an inosyl substituent. Similar results were observed for 5'-methylthioadenosine ([M+H]⁺ of 298 *m/z* and A_{max} of 260 nm) on conversion to 5'-methylthioinosine ([M+H]⁺ of 299 *m/z* and A_{max} of 250 nm) and adenosine ([M+H]⁺ of 268 *m/z* and A_{max} of 260 nm) on conversion inosine ([M+H]⁺ of 269 *m/z* and A_{max} of 250 nm) with an isobestic point at 251 nm (Supplementary Information and Supplementary Fig. 3).

Determination of SAH hydrolase activity. The homocysteinase activity of Tm0172 was determined by reaction of the free thiol group of the homocysteine product with DTNB. The increase in the absorbance at 412 nm was monitored using an extinction coefficient of 13,600 M⁻¹cm⁻¹ (ref. 37). The standard assay contained 100 mM HEPES at pH 8.0, 1.0 mM DTNB, 1.0 mM EDTA, 13 µM Tm0172 and the appropriate substrate in a final volume of 250 µl at 30 °C. The following compounds were tested for catalytic activity at concentrations up to 10 mM: S-adenosine-5'-homocysteine, S-inosyl-5'-homocysteine, 5'-methylthioadenosine and 5'-methylthioinosine. Activity was obtained only for SAH and SIH.

Data analysis. The kinetic parameters, k_{cat} , and $k_{\text{cat}}/K_{\text{m}}$ were determined by fitting the initial velocity data to the equation (1), where v is the initial velocity, E_{T} is the enzyme concentration, k_{cat} is the turnover number, S is the substrate

concentration, and K_m is the Michaelis constant³⁸. In the cases where substrate inhibition was observed an extra parameter (K_{is}) was included to calculate the apparent inhibition constant for the substrate, as observed in equation (2) (ref. 39):

$$v/E_T = k_{cat}S/(K_m + S) \quad (1)$$

$$v/E_T = k_{cat}S/[K_m + S + (S^2/K_{is})] \quad (2)$$

Sequence alignment. A multiple sequence alignment of Tm0936 with the likely orthologues from other organisms is presented in Supplementary Figure 2 of Supplementary Information.

X-ray crystallography. Tm0936 was co-crystallized with S-inosyl-homocysteine (SIH) and $ZnCl_2$. The enzyme solution at 12.9 mg ml^{-1} in 20 mM HEPES, pH 8.0 was incubated for 60 min at 4°C with 10 mM SIH and 0.5 mM $ZnCl_2$. The ternary complex was crystallized by hanging drop vapour diffusion using 1 μl of the protein–ligand solution and 1 μl of a reservoir solution containing 3.5 M Na formate, 0.5 mM $ZnCl_2$, pH 7.0. Crystals appeared in 1–2 days and exhibited diffraction consistent with the space group $P3_221$ ($a = 113.28 \text{ \AA}$, $c = 80.30 \text{ \AA}$, with 1 molecule of the ternary complex per asymmetric unit). X-ray diffraction data to 2.1 \AA were collected at the NSLS X4A beamline (Brookhaven National Laboratory) on an ADSC CCD detector. Diffraction data were integrated and scaled using the programs DENZO and SCALEPACK⁴⁰. The final 2.1 \AA data set was 93.2% complete with $R_{merge} = 0.097$.

Structure determination and model refinement. The structure of the ternary Tm0936·SIH·Zn complex was solved by molecular replacement with the program PHASER⁴¹, using apo Tm0936 (PDB code 1J6P) as the search model. The solution was subsequently refined with CNS⁴². The bound SIH and Zn were clearly visible in the electron density maps after the first cycle of rigid body refinement of the protein molecule alone. Iterative cycles of manual rebuilding with TOM⁴³ and refinement with CNS resulted in a model with R_{cryst} and R_{free} of 0.209 and 0.238, respectively. The final structure contains 3,210 protein atoms, 1 inhibitor molecule, 1 Zn atom, and 76 water molecules for one monomer of the complex in the asymmetric unit (Supplementary Information and Supplementary Table 4).

33. Irwin, J. J., Raushel, F. M. & Shoichet, B. K. Virtual screening against metalloenzymes for inhibitors and substrates. *Biochemistry* **44**, 12316–12328 (2005).
34. Gilson, M. K. & Honig, B. H. Calculation of electrostatic potentials in an enzyme active site. *Nature* **330**, 84–86 (1987).
35. Kuntz, I. D. et al. A Geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* **161**, 269–288 (1982).
36. Muszbek, L., Polgar, J. & Fesus, L. Kinetic determination of blood coagulation Factor XIII in plasma. *Clin. Chem.* **31**, 35–40 (1985).
37. Ellman, G. L. A colorimetric method for determining low concentrations of mercaptans. *Arch. Biochem. Biophys.* **74**, 443–450 (1958).
38. Cleland, W. W. Statistical analysis of enzyme kinetic data. *Methods Enzymol.* **63**, 103–138 (1979).
39. Cleland, W. W. Substrate inhibition. *Methods Enzymol.* **63**, 500–513 (1979).
40. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic Press, New York, 1997).
41. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60**, 432–438 (2004).
42. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
43. Jones, T. A. Diffraction methods for biological macromolecules. Interactive computer graphics: FRODO. *Methods Enzymol.* **115**, 157–171 (1985).

LETTERS

A turbulent wake as a tracer of 30,000 years of Mira's mass loss history

D. Christopher Martin¹, Mark Seibert², James D. Neill¹, David Schiminovich³, Karl Forster¹, R. Michael Rich⁴, Barry Y. Welsh⁵, Barry F. Madore², Jonathan M. Wheatley⁴, Patrick Morrissey¹ & Tom A. Barlow¹

Mira is one of the first variable stars ever discovered¹ and it is the prototype (and also the nearest example) of a class of low-to-intermediate-mass stars in the late stages of stellar evolution. These stars are relatively common and they return a large fraction of their original mass to the interstellar medium (ISM) (ref. 2) through a processed, dusty, molecular wind. Thus stars in Mira's stage of evolution have a direct impact on subsequent star and planet formation in their host galaxy. Previously, the only direct observation³ of the interaction between Mira-type stellar winds and the ISM was in the infrared. Here we report the discovery of an ultraviolet-emitting bow shock and turbulent wake extending over 2 degrees on the sky, arising from Mira's large space velocity and the interaction between its wind and the ISM. The wake is visible only in the far ultraviolet and is consistent with an unusual emission mechanism whereby molecular hydrogen is excited by turbulent mixing of cool molecular gas and shock-heated gas. This wind wake is a tracer of the past 30,000 years of Mira's mass-loss history and provides an excellent laboratory for studying turbulent stellar wind–ISM interactions.

Mira is a binary system with an orbital period⁴ of ~500 years. The primary star, Mira A, is the luminous, mass-shedding, evolved star that exhibits long-period, semi-regular variability. These properties arise because Mira A has exhausted the supply of hydrogen and helium in its core and at present is powered by helium and hydrogen fusion in a shell surrounding a carbon–oxygen core. These physical conditions typify what are known as asymptotic giant branch (AGB) stars. The secondary star, Mira B, is much less luminous and is usually classified as a white dwarf, but this is somewhat controversial^{5,6}. The orbital distance is so large that only a small fraction of the wind from Mira A is accreted by Mira B. Any outflow from Mira B, although potentially fast, is probably insignificant both from the perspective of mass flux and from the energetics compared to Mira A.

During routine inspection of incoming images taken with the GALEX ultraviolet satellite⁷, we noticed a nebulosity near the position of Mira. We obtained the deeper images shown in Figs 1 and 2, displaying a bow-shock-like feature south of Mira and a nebulosity extending northward in a comet-like structure spanning a total length of two degrees (see Supplementary Information for additional figures). Adopting the revised Hipparcos-based distance⁸ to Mira of 107 pc, the tail has a physical size of 4 pc. The direction of the tail is consistent with Mira's proper motion⁹ of $d(\alpha, \delta)/dt = (-28, -224) \text{ mas yr}^{-1}$ (corrected for solar motion), where α is right ascension, δ is declination and t is time. The appearance of a bow shock is consistent with Mira's large space velocity of $v_0 = 130 \text{ km s}^{-1}$, which we calculate from the proper motion and the radial velocity¹⁰ of $+63 \text{ km s}^{-1}$. Thus, the tail traces the extrapolated path of Mira over

the past 30,000 years, with the features in the tail providing an unprecedented record of Mira's wind–ISM interaction over that period. No other such ultraviolet-emitting structure is known to be connected with an AGB star. The infrared detection mentioned previously is a bow shock observed around R Hydrae³, an AGB star at a greater distance (165 pc) and with a smaller space velocity (50 km s^{-1}). However, GALEX observations of R Hydrae show no ultraviolet emission associated with the infrared bow shock.

Although Mira has been extremely well studied at all wavelengths, the bow shock and tail have previously not been detected. We outline the three major morphological features in Fig. 1: the tail, the southern bow shock, and the southern and northern 'streams' that break up into individual knots (see Fig. 3 for details). The bow shock and knots show far-ultraviolet and near-ultraviolet emission. Remarkably, the tail has only far-ultraviolet emission. The only detectable near-ultraviolet emission in the tail region is associated with the northernmost knots. Typical far-ultraviolet surface brightnesses are 30.2 ± 0.15 , 28.0 ± 0.03 and 29.6 ± 0.1 AB magnitudes per arcsec² in the bow-shock, knot and near-tail regions, respectively. These show far-ultraviolet–near-ultraviolet colours of 0.0, -0.2 and < -3 , respectively. The far-ultraviolet luminosities for the bow shock, streams and tail are 7×10^{30} , 3×10^{31} and $7 \times 10^{31} \text{ erg s}^{-1}$, respectively, for a total far-ultraviolet luminosity of $\sim 10^{32} \text{ erg s}^{-1}$.

We have searched for counterparts of these features at other wavelengths. We obtained narrow-band H α images of the nebula with the Palomar 60-inch telescope. In these images only the knots show H α , with luminosities of $L_{\text{H}\alpha} \approx 2 \times 10^{29} \text{ erg s}^{-1}$. Images of the tail place an upper limit of $L_{\text{H}\alpha} < 10^{30} \text{ erg s}^{-1}$. There are no counterparts to the tail in optical, near- or far-infrared continua. Optical spectroscopy (Fig. 3) shows evidence that the knots are shocked and ionized by the post-bow-shock flow.

These observations suggest the following physical picture. Mira A produces a cool, molecular wind¹¹; the mass loss rate is $\dot{M} \approx 3 \times 10^{-7} M_{\odot} \text{ yr}^{-1}$ (where M_{\odot} is the solar mass) and the wind velocity is $v_w \approx 5 \text{ km s}^{-1}$. The wind is mildly anisotropic and shows evidence of a bipolar component¹². The space velocity of Mira AB through the ambient interstellar medium produces a bow shock with a termination shock standoff distance $l_0 \approx 1.6 \times 10^{17} \text{ cm}$. The preshock density obtained by balancing the ram pressures at the termination shock is $n_0 \approx 0.8 \text{ cm}^{-3}$. The resulting strong shock is non-radiative, because the post-shock gas at temperature $T_1 \approx (3/16k)mv_0^2 = 5 \times 10^5 \text{ K}$ (with k being the Boltzmann constant, m being the mass of Mira's wind, and v_0 being the space velocity of Mira) and density $n_1 \approx 6 \text{ cm}^{-3}$ has an isochoric cooling time¹³ of ~2,000 years. The rise and fall of ultraviolet emissivity tracks the monotonic increase in ionization level through C^{3+} to the dominant

¹Department of Physics, Math and Astronomy, California Institute of Technology, 1200 East California Boulevard, Mail Code 405-47, Pasadena, California 91125, USA. ²Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, California 91101, USA. ³Department of Astronomy, Columbia University, New York, New York 10027, USA. ⁴Department of Physics and Astronomy, 430 Portola Plaza, UCLA, Los Angeles, California 90095-1547, USA. ⁵Space Sciences Laboratory, University of California, Berkeley, 7 Gauss Way, Berkeley, California 94720, USA.

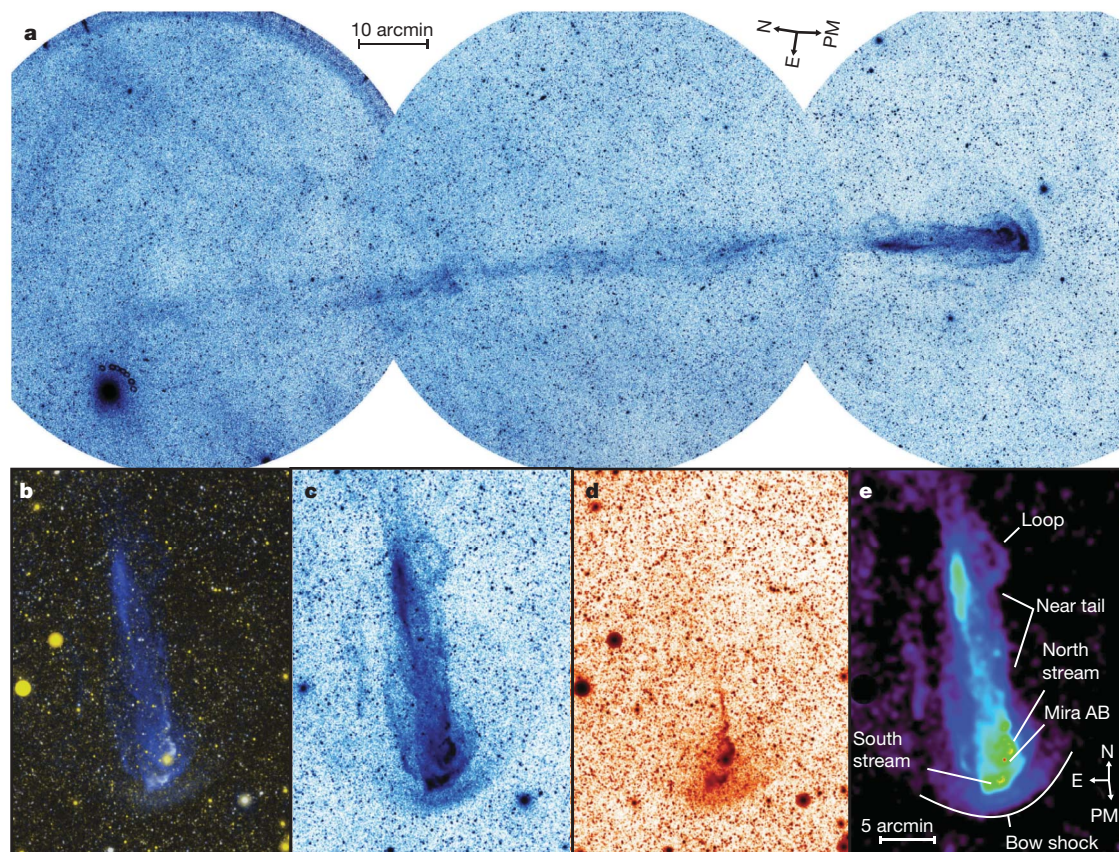


Figure 1 | Ultraviolet imaging of the Mira tail and bow shock. **a**, The far-ultraviolet (effective wavelength $\lambda_{\text{eff}} = 151.6$ nm; full-width at half-maximum, FWHM = 25.6 nm) mosaic of the Mira AB tail and bow shock. The data have been histogram equalized. The co-added mosaic consists of 31 observations covering 3 pointings with a total exposure time of 8.9, 11.2 and 11.5 ks (left to right). The images were taken between 18 November and 15 December 2006. The rectangular region is 2.7×1.1 degrees. The orientation (N, E) and measured proper motion (PM) vectors are indicated. The observed position angle of the tail is 194 degrees, in agreement with the observed proper motion vector once solar motion is accounted for. **b**, Energy-scaled colour

composite image. Far ultraviolet is blue; near ultraviolet ($\lambda_{\text{eff}} = 226.7$ nm; FWHM = 73.0 nm) is red; the average of the two ultraviolet bands is green. The region is 26×36 arcmin with an exposure time of 11.5 ks. North is up and east is left. The images were taken on 18–19 November 2006. **c**, **d**, The histogram-equalized far-ultraviolet (**c**) and near-ultraviolet (**d**) images covering the same physical region as **b**. There is a weak near-ultraviolet contribution to the bow shock and a complete lack of near-ultraviolet emission in the near-tail region. Each image has been boxcar-smoothed by 2 pixels (3 arcsec). **e**, This far-ultraviolet image has been enhanced by point source subtraction and adaptive smoothing.

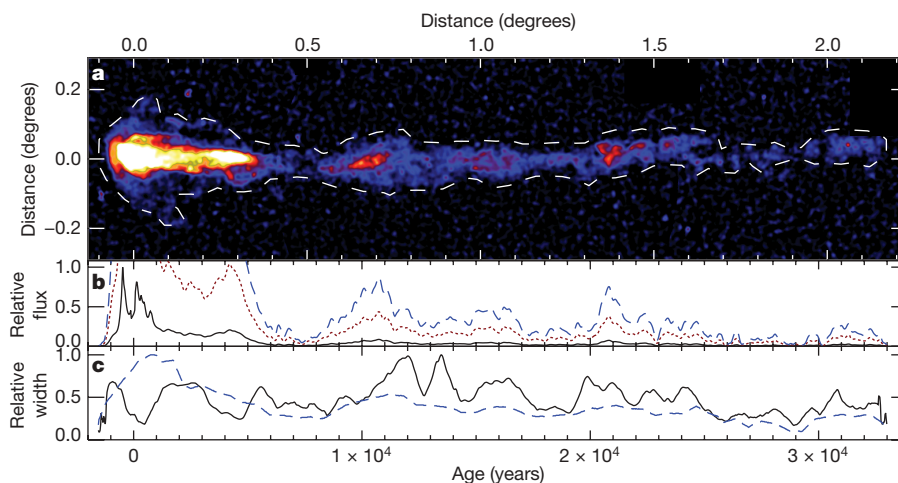


Figure 2 | 30,000-year history of Mira. **a**, Far-ultraviolet mosaic of the Mira AB tail extending more than 2 degrees. The image has been enhanced by point source subtraction and adaptive smoothing, has been rebinned by a factor of 4 into 6 arcsec pixels, and the background gradient has been subtracted. The dashed line is a total surface brightness contour of 31 AB magnitudes per arcsec². Mira AB is located at the origin (0,0). The images were taken between 18 November and 15 December 2006. **b**, Flux versus age. The solid black line is the relative flux within the surface brightness contour displayed in the top panel as a function of time, assuming a proper motion of

226 mas yr⁻¹. Zero age represents the current position of Mira AB. The dotted red line is the same multiplied by a factor of five, and the dashed blue line is the same multiplied by a factor of ten. The flux from Mira AB has been masked and is not included in the plotted flux. **c**, Width versus age. The solid black line is the relative FWHM within the surface brightness contour displayed in the top panel. The dashed blue line is the relative full width of the same contour. The time ordinate is the same as that of the middle panel. Mira AB has been masked.

C^{5+} at temperature T_1 . At the apex of the bow shock, the post-shock velocity in the Mira frame is $v_1 = 35 \text{ km s}^{-1}$. This rises with transverse position because of the conservation of the parallel velocity component across the oblique shock.

The shear flow of the hot post-shock gas past the cool wind will drag and decelerate the wind in the ISM frame. The knots have a higher density and thus more fully developed radiative shocks (Fig. 3). The position of the knots on a north–south axis is consistent with the bipolar component in Mira’s wind¹². The curvature of the southern stream of knots away from the north–south axis of the bipolar flow suggests deceleration by the post-shock flow. Knot A in the southern stream has yet to be decelerated and has the highest relative velocity. The upwind face of knot A between knot A and B is the only region showing [S II] emission. This is consistent with models of low-velocity shocks for Herbig–Haro objects¹⁴, which show a peak in [S II] emission near wind velocities of $\sim 40 \text{ km s}^{-1}$.

Most of Mira’s stellar wind is decelerated by viscous drag from the post-bow-shock flow. A central issue, therefore, is the source of the long-lived ultraviolet tail emission. It could be produced by dust scattering of interstellar light, molecular hydrogen emission, the hydrogen two-photon continuum, or coronal line emission from shocks or intermediate-temperature gas at $\sim 10^5 \text{ K}$. We have used

GALEX grism mode observations to constrain the spectrum of the emission. The details of this analysis are provided in Supplementary Information. Although AGB winds are dusty, dust-scattered far-ultraviolet emission is ruled out by a poor grism fit, the lack of observed near-ultraviolet emission, and the brightness of the emission in the absence of an illumination source other than the interstellar radiation field. The process providing the best fit to the grism image is H_2 emission. It also produces no detectable counterparts at other wavelengths. The far-ultraviolet emission cannot be fluorescence of H_2 excited by the ambient interstellar ultraviolet radiation field¹⁵ because this fails to produce the observed brightness by more than a factor of seven. Rather, we propose that the emission is excited collisionally¹⁶ by the interaction of H_2 in the cool wind wake with hot electrons in the post-shock gas resulting from the bow shock that also entrains and decelerates the wind. This is in accordance with the fact that the ultraviolet-emitting gas is likely to be cool, because the pressure of a hot gas would be much higher than the ambient ISM making it hard to maintain the narrow collimation, observed in Figs 1 and 2, for 30,000 years.

Indeed, this emission mechanism provides consistency between the global properties of the nebula and the transition rates expected from collisional excitation, and is consistent with a near equilibrium

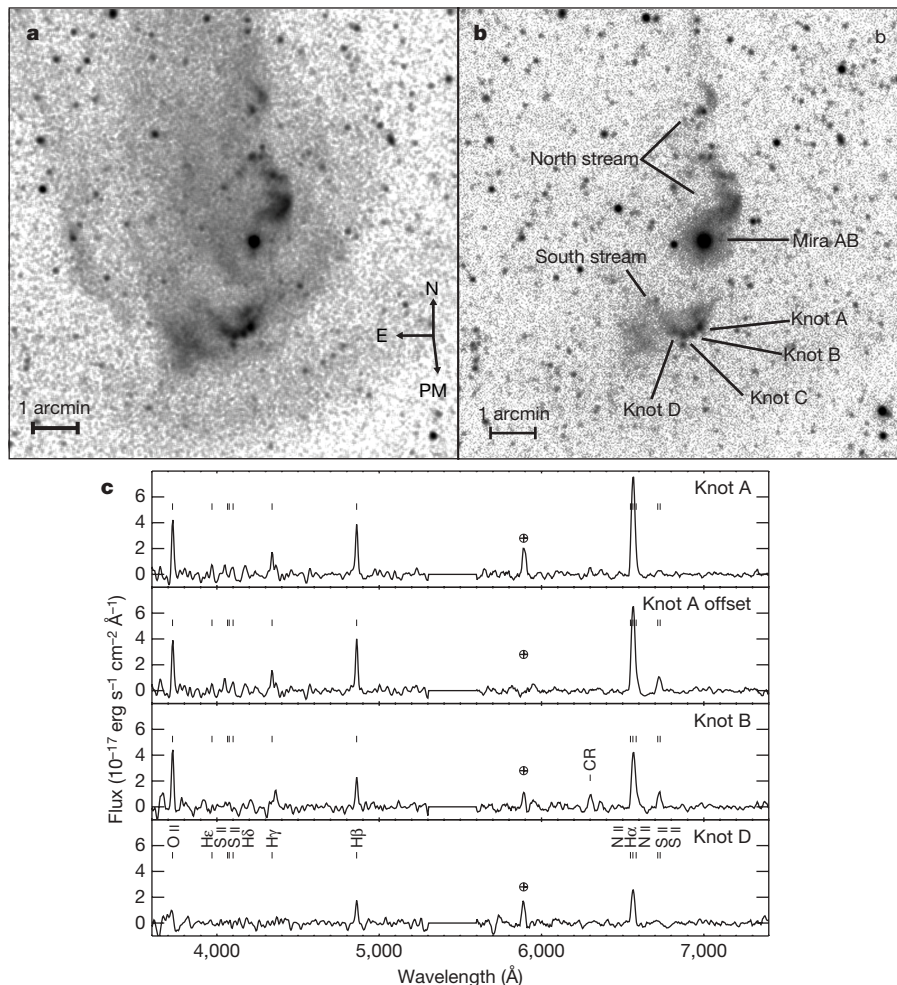


Figure 3 | Ultraviolet images and optical spectra of the streams and knots. **a, b**, The far-ultraviolet (**a**) and near-ultraviolet (**b**) images of the $10 \times 10 \text{ arcmin}^2$ region around Mira AB. North is up and east is left. **c**, The spectra were taken on 12 December 2006 with the Double Spectrograph on the Palomar 5-m telescope. The extractions were centred on the knots indicated, with the exception of the knot A offset, which was moved to coincide with the [S II] emission between knots A and B. The presence of the Balmer lines and the [O II] and [S II] lines is consistent with a low-velocity

secondary shock formed around each knot because the post-bow-shock flow affects the knots with a velocity of 35 km s^{-1} . This flow is slowing down the upstream velocity of each knot successively, thus causing the arc in their positions. The [S II] appearing at the leading edge of knot A is consistent with this picture because knot A would have the largest relative velocity with respect to the flow. The circled cross indicates night sky lines and CR indicates cosmic ray hits.

between dissociation and replenishment of H_2 from Mira's wind. At the interface of cool wind and hot post-shock gas, ~ 30 -eV electrons from the latter excite H_2 in the former into the $\text{B}^1\Sigma_g^+$ (Lyman) electronic level. The tail far-ultraviolet luminosity implies 5×10^{42} molecules s^{-1} are radiating. On the basis of the collisional excitation cross-sections into unbound states¹⁷, the dissociation rate will be half this, in rough equilibrium with the H_2 replenishment rate of $\sim 3 \times 10^{42}$ molecules s^{-1} from the predominantly molecular¹⁸ AGB wind with $\dot{M} \approx 3 \times 10^{-7} M_\odot \text{yr}^{-1}$. We assume that the average wind flux over the past $\sim 10^4$ years is similar to that measured at present. This ratio of photon emission to dissociation has also been observed in Herbig–Haro objects¹⁹. Collisional excitation, emission and dissociation would occur in the thin multiphase layer ($\sim 10^{15}$ – 10^{16} cm), with the total rate being the product of the electron velocity, electron density and interface area ($v_e n_e A$). Using the post-shock temperature and density, the implied interface area is $\sim (5 \times 10^{16})^2 \text{cm}^2$, consistent with the nebular dimensions. Emission from a multiphase surface can produce a more complex morphology in the tail than volume emission, as is observed.

The total mechanical energy available to power the nebula is estimated by assuming that the local ISM brings the wind (again at the current mass flux) to rest from its space velocity of 130 km s^{-1} . The result, $2 \times 10^{33} \text{ erg s}^{-1}$, is 20 times the observed far-ultraviolet luminosity.

Mira A is a thermally pulsing AGB star, and the wind flux may vary over the thermal pulse cycle²⁰. Signatures of thermal pulsation in AGB winds have previously been observed in the infrared²¹, but never before in the ultraviolet. Figure 2 illustrates the evidence for a cyclic behaviour in Mira's wind, with maxima falling roughly every $\sim 10^4$ years. A periodogram²² shows a highly significant peak around 10^4 years and a secondary peak at approximately half this value. The bottom panel of Fig. 2, plotting the wake width versus time, provides additional evidence for this periodic timescale. The standoff distance and characteristic radius of the wake, l_0 , vary with wind flux consistently with the expected $l_0 \propto \dot{M}^{1/2}$.

We can compare this timescale with that predicted from the thermal pulse theory. Mira's mass, derived from the well established period–mass–radius relationship²³ for long-period variable stars, is $\sim 1.5 M_\odot$. This mass is consistent with models predicting a thermal pulse cycle at least an order of magnitude longer²⁰ than the periodicity timescale in Mira's wake. The only thermal pulse models consistent with the observed wind wake periodic timescale have masses of $\sim 4 M_\odot$ (ref. 20). Alternatively, the periodic features could be density fluctuations caused by large-scale turbulence. Mira's cool wind has a high Reynolds number ($\sim 10^4$), and the wind wake may grow turbulent. The large loop on the west side of the tail could be a decaying turbulent vortex, as predicted by hydrodynamic simulations of AGB winds interacting with the ISM^{24,25}. Variations in the ISM density on parsec scales could provide another alternative explanation.

A deeper understanding of the Mira nebula requires detailed hydrodynamic and microphysical modelling. Ultra-deep optical and near-infrared observations may reveal the expected extremely faint H_2 emission lines, whereas far-infrared observations may uncover warm, entrained dust in the wake. Absorption-line observations of background stars in the H_2 Lyman and Werner bands would be extremely interesting, as would observations of low-ionization atomic species. The discovery of a two-degree-long wind wake emitting only in the far ultraviolet provides an unprecedented fossil record of post-main-sequence stellar evolution and mass loss, a laboratory for the study of astrophysical turbulence and the complex

physics of a multiphase hydrodynamical flow, and suggests a new cooling process for hot gas that entrains a cool molecular phase. After 400 years of study, Mira continues to astound.

Received 30 March; accepted 11 June 2007.

- Hoffleit, D. History of the discovery of Mira stars. *J. Am. Ass. Var. Star Observ.* **25**, 115–136 (1997).
- Schröder, K.-P. & Sedlmayr, E. The galactic mass injection from cool stellar winds of the 1 to $2.5 M_\odot$ stars in the solar neighbourhood. *Astron. Astrophys.* **366**, 913–922 (2001).
- Ueta, T. S. et al. Detection of a far-infrared bow shock nebula around R Hya: the first MIRIAD results. *Astrophys. J.* **648**, L39–L42 (2006).
- Prieur, J. L. et al. High angular resolution observations of late-type stars. *Astrophys. J.* **139**, 249–258 (2002).
- Karovska, M., Schlegel, E., Hack, W., Raymond, J. C. & Wood, B. E. A. Large X-ray outburst in Mira A. *Astrophys. J.* **623**, L137–L140 (2005).
- Ireland, M. J. et al. Born again protoplanetary disk around Mira B. *Astrophys. J.* **662**, 651–657 (2007).
- Martin, D. C. et al. The Galaxy Evolution Explorer: a space ultraviolet survey mission. *Astrophys. J.* **619**, L1–L4 (2005).
- Knapp, G. R., Pourbaix, D., Platais, I. & Jorissen, A. Reprocessing the Hipparcos data of evolved stars. III. Revised Hipparcos period–luminosity relationship for galactic long-period variable stars. *Astron. Astrophys.* **403**, 993–1002 (2003).
- Turon, C. et al. Version 2 of the Hipparcos Input Catalogue. *Bull. Inf. Centre Données Stell.* **43**, 5–6 (1993).
- Evans, D. S. The revision of the general catalog of radial velocities. *IAU Symp.* **30**, 57–62 (1967).
- Ryde, N., Gustafsson, B., Eriksson, K. & Hinkle, K. H. Mira's wind explored in scattering infrared CO lines. *Astrophys. J.* **545**, 945–956 (2000).
- Josselin, E. et al. Strong asymmetries in the neutral envelope of Mira. *Astron. Astrophys.* **362**, 255–262 (2000).
- Ferland, G. J. et al. CLOUDY 90: numerical simulation of plasmas and their spectra. *Publ. Astron. Soc. Pacif.* **110**, 761–778 (1998).
- Hartigan, P., Raymond, J. & Hartmann, L. Radiative bow shock models of Herbig–Haro objects. *Astrophys. J.* **316**, 323–348 (1987).
- Martin, C., Hurwitz, M. & Bowyer, S. Discovery of molecular hydrogen fluorescence in the diffuse interstellar medium. *Astrophys. J.* **354**, 220–228 (1990).
- Liu, W. & Dalgarno, A. The ultraviolet spectrum of the Jovian dayglow. *Astrophys. J.* **462**, 502–518 (1996).
- Khakoo, M. A., Trajimar, S., McAdams, R. & Shyn, T. W. Electron-impact excitation cross sections for the $b^3 \sum_u^+$ state of H_2 . *Phys. Rev. A* **35**, 2832–2837 (1987).
- Bowers, P. F. & Knapp, G. R. Detection of H I emission in the circumstellar envelope of Omicron Ceti (Mira). *Astrophys. J.* **332**, 299–304 (1988).
- Raymond, J. C., Blair, W. P. & Long, K. S. Hopkins ultraviolet telescope observations of H_2 emission lines from HH 2. *Astrophys. J.* **489**, 314–318 (1997).
- Vassiliadis, E. & Wood, P. R. Evolution of low- and intermediate-mass stars to the end of the asymptotic giant branch with mass loss. *Astrophys. J.* **413**, 641–657 (1993).
- Speck, A. K., Meixner, M. & Knapp, G. R. Discovery of parsec-sized dust shells around AFGL 2688 and AFGL 618. *Astrophys. J.* **545**, L145–L148 (2000).
- Scargle, J. D. Studies in astronomical time series analysis. II—Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263**, 835–853 (1982).
- Wood, P. R. in *Miras to Planetary Nebulae: Which Path for Stellar Evolution?* (eds Mennessier, M. O. & Omont, A.) 67–84 (Editions Frontières, Gif-sur-Yvette, 1990).
- Wareing, C. J. et al. Detached shells as tracers of asymptotic giant branch–interstellar medium bow shocks. *Mon. Not. R. Astron. Soc.* **372**, L63–L67 (2006).
- Wareing, C. J., Zijlstra, A. A. & O'Brien, T. J. Vortices in the wakes of AGB stars. *Astrophys. J.* **660**, L129–L132 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Cenko, S. Browne, S. Kulkarni and F. Harrison for assistance in obtaining optical data, and M. Shara and P. Szkody for comments. This work was supported by the National Aeronautics and Space Administration.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.C.M. (cmartin@srl.caltech.edu).

LETTERS

Generation of optical ‘Schrödinger cats’ from photon number states

Alexei Ourjoumtsev¹, Hyunseok Jeong², Rosa Tualle-Brouri¹ & Philippe Grangier¹

Schrödinger’s cat¹ is a *Gedankenexperiment* in quantum physics, in which an atomic decay triggers the death of the cat. Because quantum physics allow atoms to remain in superpositions of states, the classical cat would then be simultaneously dead and alive. By analogy, a ‘cat’ state of freely propagating light can be defined as a quantum superposition of well separated quasi-classical states^{2,3}—it is a classical light wave that simultaneously possesses two opposite phases. Such states play an important role in fundamental tests of quantum theory^{4–7} and in many quantum information processing tasks, including quantum computation⁸, quantum teleportation^{9,10} and precision measurements¹¹. Recently, optical Schrödinger ‘kittens’ were prepared^{12–14}; however, they are too small for most of the aforementioned applications and increasing their size is experimentally challenging. Here we demonstrate, theoretically and experimentally, a protocol that allows the generation of arbitrarily large squeezed Schrödinger cat states, using homodyne detection and photon number states as resources. We implemented this protocol with light pulses containing two photons, producing a squeezed Schrödinger cat state with a negative Wigner function. This state clearly exhibits several quantum phase-space interference fringes between the ‘dead’ and ‘alive’ components, and is large enough to become useful for quantum information processing and experimental tests of quantum theory.

The predictions of quantum physics for microscopic objects cannot be simply generalized to our ‘classical’ world. In fact, the reason why Schrödinger’s cats are so hard to prepare is the same that makes large quantum computers so hard to build: for macroscopic systems, quantum state superpositions rapidly decohere into statistical mixtures because of strong interactions with the environment. To become feasible, Schrödinger’s *Gedankenexperiment* should be transposed from a cat to a more convenient physical system, with its own ‘classical’ or ‘quasi-classical’ states. In quantum optics, they correspond to coherent states $|\alpha\rangle$, where α is the coherent amplitude^{15,16}. Therefore, a quantum superposition $|\psi\rangle = \mathcal{N}(|\alpha\rangle + e^{i\theta}|- \alpha\rangle)$ defines a optical cat state with a ‘size’ $|\alpha|^2$, where $\mathcal{N} = \left[2\left(1 + \cos(\theta)e^{-2|\alpha|^2}\right)\right]^{-1/2}$ is a normalization constant. As the phase origin is arbitrary, we will assume in the following that α is real.

In addition to their numerous applications^{4–11}, optical cat states have another crucial advantage: quantum optics provide efficient tools to tell the difference between a true quantum superposition and a plain statistical mixture of two coherent states. Quantum states of light, often considered in terms of photons, can also be described as waves, using their amplitudes and phases or, in cartesian coordinates, their quadratures \hat{x} and \hat{p} (ref. 17). A state is then characterized by the quasi-probability distribution of its quadratures $W(x, p)$, called the Wigner function¹⁸. It can be reconstructed by homodyne tomography from several marginal quadrature distributions $P(\hat{x}_\theta = \hat{x} \cos \theta + \hat{p} \sin \theta)$ measured with a homodyne detection. As \hat{x}

and \hat{p} are not simultaneously defined in quantum physics, the Wigner function may become negative for specific quantum states, including optical ‘Schrödinger cats’. In this case, the Wigner function clearly reveals the difference between a real quantum superposition and a mere statistical mixture of the two states $|\pm\alpha\rangle$: for a true superposition state, it presents a phase-space interference between the ‘dead’ ($|- \alpha\rangle$) and ‘alive’ ($|\alpha\rangle$) components and takes negative values.

Such superposition states could only be observed in bound systems^{19,20} until, very recently, several groups succeeded in preparing free-propagating light beams in small cat states (‘Schrödinger kittens’)^{12–14}. These experiments attracted much attention, as first steps on a new promising way towards quantum communication. But the size of the ‘kittens’ accessible so far is limited to $|\alpha|^2 \lesssim 1$, and their amplification²¹ remains a serious experimental challenge²². On the other hand, most quantum information processing applications require larger cats with a smaller overlap between the two coherent states: it should be typically less than 1%, which corresponds to $|\alpha|^2 \gtrsim 2.3$.

In this letter we demonstrate, theoretically and experimentally, a method to produce quantum superpositions of squeezed coherent states with arbitrarily large amplitudes (see Fig. 1). These cat states are squeezed along the x quadrature and stretched along p , which makes them more robust against decoherence²³. If needed, they can be easily ‘un-squeezed’, either by injecting them into a degenerate optical parametric amplifier, or by mixing them with squeezed vacuum^{24,25}. The required squeezing, around 3 dB, is easily achievable.

The basis of our protocol is to split a photon number state (Fock state) containing exactly n photons on a 50/50 beam splitter (BS), and to measure the momentum quadrature \hat{p} in one mode. The desired state is prepared in the other mode, under the condition that $|p| \leq \varepsilon \ll 1$.

An interesting insight into the structure of the prepared state is obtained by looking at its wavefunction ϕ_m in the limit $\varepsilon \rightarrow 0$ (as we show below, a finite ε is a second-order effect that does not perturb our experiments). Omitting the normalization factors, the wave function of a n -photon number state in the momentum quadrature basis is $H_n(p)e^{-p^2/2}$, where H_n is the n th Hermite polynomial. Mixed

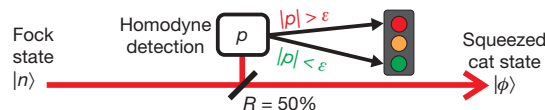


Figure 1 | Preparing squeezed ‘Schrödinger cat’ states from Fock states using a single homodyne detection. A photon number state containing n photons is divided into two modes on a beam splitter with 50% reflectivity. A homodyne detection measures the momentum quadrature \hat{p} in one mode. If the measurement outcome p is close to 0 within an acceptance width ε ($|p| \leq \varepsilon \ll 1$), the other mode is successfully prepared in a ‘squeezed cat’ state, otherwise it is discarded. See text for details.

¹Laboratoire Charles Fabry de l’Institut d’Optique, Université Paris-Sud, CNRS UMR 8501, 91127 Palaiseau, France. ²Centre for Quantum Computer Technology, Department of Physics, University of Queensland, Brisbane, Queensland 4072, Australia.

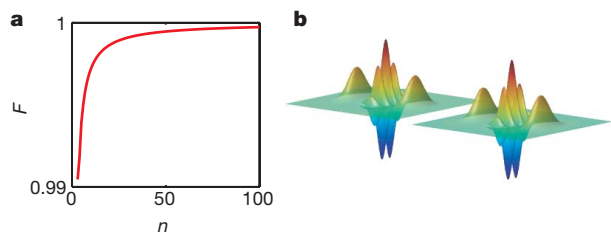


Figure 2 | Theoretical performance. **a**, Fidelity F between the state produced with n photons and an ideal Schrödinger cat with a 'size' $|\alpha|^2 = n$, squeezed by 3 dB. **b**, Example of ideal state preparation. The Wigner function of the pure state prepared from 10 photons (left) compared with an ideal Schrödinger cat state with $\alpha = \sqrt{10}$ squeezed by 3 dB (right). Their fidelity is $F_{10} \approx 99.7\%$.

with vacuum on a 50/50 BS, the two-mode wavefunction becomes $\tilde{\phi}(p, p_0) = H_n[(p - p_0)/\sqrt{2}]e^{-(p^2 + p_0^2)/2}$. If the measurement outcome is $p_0 = 0$, by taking the Fourier transform we see that the un-normalized wavefunction in the position basis is simply $\phi_n(x) = x^n e^{-x^2/2}$.

For $n \geq 3$, this state has a fidelity $F_n > 99\%$ with a 'Schrödinger cat' state with a size $|\alpha|^2 = n$ and a superposition phase $\theta = n\pi$, which has been squeezed by 3 dB along the x axis. Remarkably enough, the quality of the prepared 'cats' increases with their size, as shown in Fig. 2a. We see numerically that $F_n \approx 1 - 0.03/n$, and we rigorously prove in the Supplementary Information that the fidelity tends to 1 when $n \rightarrow \infty$. For small n we observe a slight deviation from this scaling law: when $n = 2$, a cat state with $|\alpha|^2 = 2.6$ squeezed by 3.5 dB is obtained with a 99% fidelity. As an illustration, we present in Fig. 2b the Wigner function of the pure state prepared with 10 photons, compared to an ideal cat state $\mathcal{N}(|\sqrt{10}\rangle + |-\sqrt{10}\rangle)$ squeezed by 3 dB. In this case the fidelity is $F_{10} \approx 99.7\%$.

We implemented this protocol experimentally using ultrashort light pulses (180 fs) prepared in $n = 2$ number states (see Fig. 3). Their preparation is detailed elsewhere²⁶. In brief, two beams containing the same number of photons (two-mode squeezed state) are produced in a spatially non-degenerate optical parametric amplifier

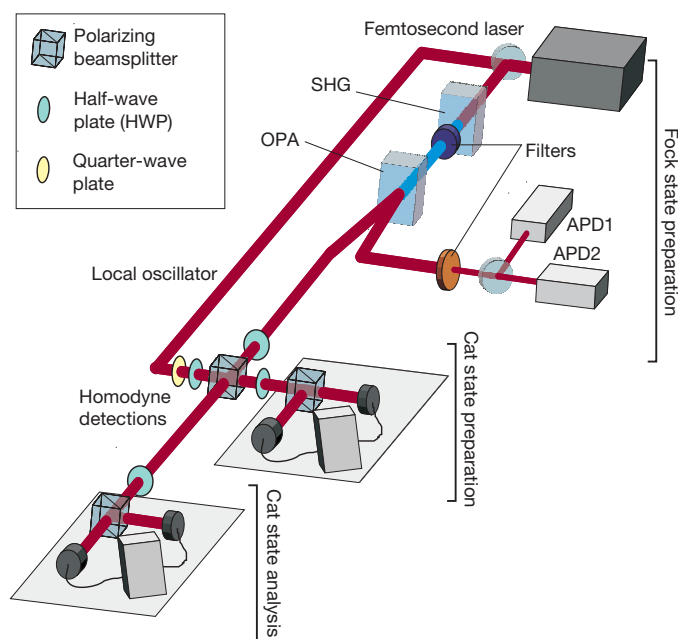


Figure 3 | Experimental set-up. Femtosecond pulses, frequency-doubled by second harmonic generation (SHG), pump a spatially degenerate optical parametric amplifier (OPA). A two-photon state is prepared in one mode by a coincidence detection in the other, using two avalanche photodiodes (APD). It is split between two homodyne detectors: one is used for the preparation of the cat state, the other for the analysis (see text for details).

(OPA) by down-conversion of frequency-doubled femtosecond laser pulses. One of them is split between two avalanche photodiodes (APDs) after spatial and spectral filtering. A coincidence APD detection heralds the presence of at least two photons, and as the parametric gain is not too large ($g = 1.17$), this projects the other mode in a two-photon number state.

These $n = 2$ states are split on a 50/50 BS. The reflected mode is measured by a time-resolved homodyne detection. We accept the outcome p if $|p| < 0.1$, which leads to a success probability of $\sim 7.5\%$. This prepares the desired 'squeezed cat' states in the other mode, with a rate of $\sim 7 \text{ s}^{-1}$. To analyse these states, we perform a homodyne tomography with a second detection, measuring six different quadrature distributions with 15,000 data points each. From these distributions, using a maximal-likelihood algorithm, we reconstruct the Wigner function of the prepared state corrected for the losses of the final homodyne detection. We note that the defects of the first detection, involved in the preparation of the state, cannot be compensated.

The reconstructed Wigner function, presented in Fig. 4, is clearly negative. We observe the expected phase-space interference between two coherent states with amplitudes $\alpha = \pm \sqrt{2.6}$ squeezed by 3.5 dB. As shown below, the difference from the ideal 'squeezed cat' is essentially due to technical issues.

The prepared states are very sensitive to experimental imperfections. Dark counts and stray light decrease the probability ξ for an APD detection to correspond to the desired photon number state preparation. Mode distortion in the nonlinear crystals and imperfect laser beams lead to an impure initial two-mode squeezed state. We can consider that the associated excess noise is added by phase-independent parametric amplification with a gain $h = \cosh(\gamma r)^2$ on

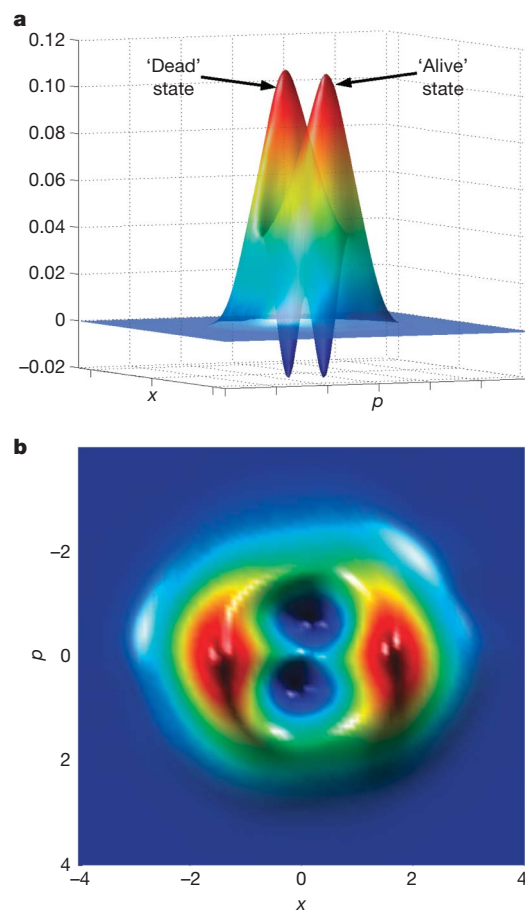


Figure 4 | Experimental results. **a**, **b**, Experimental Wigner function $W(x, p)$ produced with $n = 2$ photons, corrected for the losses of the final homodyne detection (**a**, side view; **b**, top view). An interference between the 'dead' and 'alive' states with two negative regions is clearly visible.

a pure two-mode state squeezed by $s = \exp(-2r)$, where γ is the relative efficiency of the amplification process responsible for the excess noise. The limited efficiency η and the noise ϵ of the homodyne detection involved in the state preparation decrease the purity of the final state. The defects of the second detection are not involved in the preparation but only in the analysis of the generated states, and must be corrected for to determine the actual Wigner function.

Taking all these parameters into account, we derived an analytical model for the generated states (see Supplementary Information). Figure 5a presents the Wigner function obtained from our model with the actual values of the experimental parameters, and we see it is extremely similar to the one reconstructed from the data with the maximal-likelihood algorithm. Figure 5b shows the Wigner function we would obtain with pure photon number states and a lossless detection for the same acceptance width $\epsilon = 0.1$, compared to the pure case ($\epsilon = 0$). Their fidelity is 99%, which shows that our experiment is limited by technical issues and not by the finite ϵ . In a general case, the effect of ϵ is discussed in Supplementary Information.

Obviously, the states produced with this protocol contain at most n photons. For example, for $n = 2$, $|\phi_2\rangle = \sqrt{2/3}|2\rangle - \sqrt{1/3}|0\rangle$. It is quite easy to see that we can prepare ‘even’ or ‘odd’ cat states, containing only even or only odd photon numbers, depending on the parity of n . Indeed, in this case the homodyne detection performs a parity measurement: an outcome $p = 0$ tells us that the number of photons in the measured mode was even, as the overlap between the (non-physical) state $|p = 0\rangle$ and an odd photon number state is null. Therefore, the prepared state has the same parity as n . The homodyne measurement also induces a phase dependence on the originally phase-invariant state. The 3 dB of squeezing required to ‘unsqueeze’ a cat with $|\alpha|^2 = n$ compensate for the loss of half of the photons on the 50/50 BS without changing the parity. Another way to intuitively deduce the scaling law $|\alpha|^2 = n$ is to notice that $\phi_n(x)$ presents two peaks centred at $x = \pm\sqrt{n}$, whereas for an ideal cat squeezed by $s_c = 1/2$ (3 dB) the same two peaks are at $x = \pm\sqrt{2s_c}|\alpha|^2$. In fact, we can also prepare a cat with a slightly different size by changing s_c if we preserve the relationship $n = 2s_c|\alpha|^2$. For example, a odd cat with $|\alpha|^2 = 9.5$ can be prepared with a 99.7% fidelity using 9 photons with $s_c = 9/19$. For an even cat we would use 10 photons and ‘unsqueeze’ by $s_c = 10/19$. Therefore, all parities and cat sizes are accessible.

We have thus proposed and experimentally demonstrated a protocol that allows the preparation of quantum superpositions of squeezed coherent states. Considering the fast technical progress and the increasing number of groups working in this field, we expect that

the purity of these superpositions will rapidly improve. The use of higher parametric gains combined with number-resolving photon counters allows the preparation of higher photon number states²⁷, and should give access to even larger ‘Schrödinger cats’. This simple and flexible procedure is particularly suitable for producing these states as ‘ancillas’ for numerous quantum information processing tasks.

Received 11 April; accepted 26 June 2007.

- Schrödinger, E. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* **23**, 807–812 (1935).
- Yurke, B. & Stoler, D. Generating quantum mechanical superpositions of macroscopically distinguishable states via amplitude dispersion. *Phys. Rev. Lett.* **57**, 13–16 (1986).
- Schleich, W., Pernigo, M. & Kien, F. L. Nonclassical state from two pseudoclassical states. *Phys. Rev. A* **44**, 2172–2187 (1991).
- Sanders, B. C. Entangled coherent states. *Phys. Rev. A* **45**, 6811–6815 (1992).
- Wenger, J., Hafezi, M., Grosshans, F., Tualle-Broui, R. & Grangier, P. Maximal violation of Bell inequalities using continuous-variable measurements. *Phys. Rev. A* **67**, 012105 (2003).
- Jeong, H., Son, W., Kim, M. S., Ahn, D. & Brukner, Č. Quantum nonlocality test for continuous-variable states with dichotomic observables. *Phys. Rev. A* **67**, 012106 (2003).
- Stobińska, M., Jeong, H. & Ralph, T. C. Violation of Bell’s inequality using classical measurements and nonlinear local operations. *Phys. Rev. A* **75**, 052105 (2007).
- Ralph, T. C. et al. Quantum computation with optical coherent states. *Phys. Rev. A* **68**, 042319 (2003).
- Van Enk, S. J. & Hirota, O. Entangled coherent states: Teleportation and decoherence. *Phys. Rev. A* **64**, 022313 (2001).
- Jeong, H., Kim, M. S. & Lee, J. Quantum-information processing for a coherent superposition state via a mixed entangled coherent channel. *Phys. Rev. A* **64**, 052308 (2001).
- Munro, W. J., Nemoto, K., Milburn, G. J. & Braunstein, S. L. Weak-force detection with superposed coherent states. *Phys. Rev. A* **66**, 023819 (2002).
- Ourjoumtsev, A., Tualle-Broui, R., Laurat, J. & Grangier, P. Generating optical Schrödinger kittens for quantum information processing. *Science* **312**, 83–86 (2006).
- Neergaard-Nielsen, J. S., Nielsen, B. M., Hettich, C., Mølmer, K. & Polzik, E. S. Generation of a superposition of odd photon number states for quantum information networks. *Phys. Rev. Lett.* **97**, 083604 (2006).
- Wakui, K., Takahashi, H., Furusawa, A. & Sasaki, M. Controllable generation of highly nonclassical states from nearly pure squeezed vacua. *Opt. Express* **15**, 3568–3574 (2007).
- Schrödinger, E. Der stetige Uebergang von der Mikro- zur Makromechanik. *Naturwissenschaften* **14**, 664–666 (1926).
- Zurek, W. H., Habib, S. & Paz, J. P. Coherent states via decoherence. *Phys. Rev. Lett.* **70**, 1187–1190 (1993).
- Walls, D. F. & Milburn, G. J. *Quantum Optics* (Springer, Berlin, 1994).
- Wigner, E. P. On the quantum correction for thermodynamic equilibrium. *Phys. Rev.* **40**, 749–759 (1932).
- Brune, M. et al. Observing the progressive decoherence of the “meter” in a quantum measurement. *Phys. Rev. Lett.* **77**, 4887–4890 (1996).
- Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. A. “Schrödinger cat” superposition state of an atom. *Science* **272**, 1131–1135 (1996).
- Lund, A. P., Jeong, H., Ralph, T. C. & Kim, M. S. Conditional production of superpositions of coherent states with inefficient photon detection. *Phys. Rev. A* **70**, 020101(R) (2004).
- Suzuki, S., Takeoka, M., Sasaki, M., Andersen, U. & Kannari, F. Practical purification scheme for decohered coherent-state superpositions via partial homodyne detection. *Phys. Rev. A* **73**, 042304 (2006).
- Serafini, A., De Siena, S., Illuminati, F. & Paris, M. G. A. Minimum decoherence cat-like states in Gaussian noisy channels. *J. Opt. B* **6**, S591–S596 (2004).
- Jeong, H. et al. Quantum-state engineering with continuous-variable postselection. *Phys. Rev. A* **74**, 033813 (2006).
- Yoshikawa, J. et al. Demonstration of a high-fidelity, deterministic and universal squeezing transformation. Preprint at (<http://arxiv.org/quant-ph/0702049>) (2007).
- Ourjoumtsev, A., Tualle-Broui, R. & Grangier, P. Quantum homodyne tomography of a two-photon number state. *Phys. Rev. Lett.* **96**, 213601 (2006).
- Waks, E., Diamanti, E. & Yamamoto, Y. Generation of photon number states. *N. J. Phys.* **8**, 4 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported (in France) by the EU IST/FET project COVAQIAL and the ANR/PNANO project IRCOQ, and (in Australia) by the US Army Research Office and the DTO, the Australian Research Council and Queensland State Government. H.J. thanks T. C. Ralph and M. S. Kim for discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.O. (alexei.ourjoumtsev@institutoptique.fr).

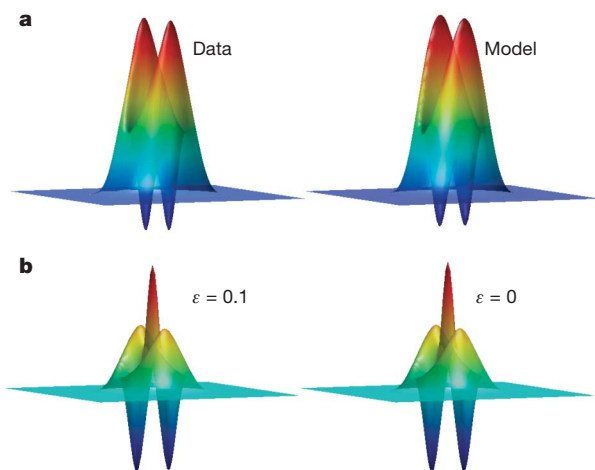


Figure 5 | Influence of experimental imperfections. **a**, Experimental Wigner function (left) produced from two-photon Fock states, compared to that obtained with our model (right). **b**, Wigner function produced from pure two-photon Fock states with $\epsilon = 0.1$ and no technical imperfections (left) compared to the ideal case with $\epsilon = 0$ (right).

Vitrification of a monatomic metallic liquid

M. H. Bhat¹, V. Molinero^{1,2}, E. Soignard¹, V. C. Solomon¹, S. Sastry³, J. L. Yarger¹ & C. A. Angell¹

Although the majority of glasses in use in technology are complex mixtures of oxides or chalcogenides, there are numerous examples of pure substances—‘glassformers’—that also fail to crystallize during cooling. Most glassformers are organic molecular systems, but there are important inorganic examples too^{1,2}, such as silicon dioxide and elemental selenium (the latter being polymeric). Bulk metallic glasses can now be made³; but, with the exception of Zr₅₀Cu₅₀ (ref. 4), they require multiple components to avoid crystallization during normal liquid cooling. Two-component ‘met-glasses’ can often be achieved by hyperquenching, but this has not hitherto been achieved with a single-component system. Glasses form when crystal nucleation rates are slow, although the factors that create the slow nucleation conditions are not well understood. Here we apply the insights gained in a recent molecular dynamics simulation study⁵ to create conditions for successful vitrification of metallic liquid germanium. Our results also provide micro-graphic evidence for a rare polyamorphic transition preceding crystallization of the diamond cubic phase.

It has been argued that any liquid can be vitrified if the cooling rate is sufficiently high⁶, and even ideal gases can be vitrified if the molecules are allowed complex shapes⁷. It seems that the crystal nucleation time can be made long with respect to amorphous form assembly times in a variety of ways, for instance by making sure that not more than a few atoms are hot at any time. Thus amorphous forms of metals like Bi and Pb have been made by deposition from the vapour at very low temperatures⁸. However, these prove to have no kinetic stability, and crystallize when the temperature is raised even 15 K above absolute zero. In spite of the conclusions of ref. 6, vitrification of single-component metals by cooling from the liquid might therefore seem improbable.

On the other hand, we demonstrated recently⁵ by molecular dynamics that an atomic liquid, Stillinger–Weber silicon, becomes non-crystallizing in ‘slow’ molecular dynamics cooling runs when the interaction potential is modified, without reducing the attractive potential, so as to lower the melting point by ~50% and put isoenergetic crystals in competition with each other against increasingly stabilized liquid. We found that ability to vitrify on the computational timescale was established when the diffusivity of the liquid at the crystallization temperature was reduced to $1.0 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$, and the excess free energy, which provides the drive to crystallize below the melting temperature T_m , rose most slowly during super-cooling. These turn out to be the same conditions established⁹ in experiments on binary alloys such as Ni–P (ref. 10) and Zr₂Ni, which are vitrifiable by melt spinning. (In these cases, the melting point of Ni is lowered by 35%, and Zr by 42%, as a result of chemical mixing in which an optimum negative deviation from ideal mixing is exploited^{3,6}. In Zr–Cu, a ‘bulk’ glassformer at 1:1, a metastable eutectic⁴ lies lower still).

These results suggested to us that another variable, pressure, might be used to achieve the same conditions for a single-component metal of the right initial properties. Pressure can only lower the melting

point if the melting is accompanied by a volume decrease, so the possibilities, starting at zero pressure, are limited to Bi, Ga, Ce, Si and Ge. Having used liquid Si as the starting point in our ‘potential tuning’ study⁵, it was natural to choose liquid Si for initial tests of the new proposal, using molecular dynamics simulation of the Stillinger–Weber silicon potential¹¹.

At the solid–solid–liquid triple point T_3 of Si, T_3/T_m (1 atm) is 0.5 (see Methods and Supplementary Information), and simulations with the Stillinger–Weber potential have found¹² (see Methods and Supplementary Information) that its diffusivity at T_3 was $0.9 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$, as in the Ni–P eutectic. Encouraged by these relations and also by Brazhkin’s observations on In–Sb (ref. 13), and Tanaka’s arguments¹⁴, but having failed to vitrify Si with multianvil cell quenching at 9 GPa (see Supplementary Information), we turned our attention to experiments conducted in a high-pressure diamond anvil cell (DAC).

In DACs, the transparency of the compressing diamonds permits the use of pulsed-laser-beam melting methods. The high thermal conductivity of the diamonds ensures rapid cooling of the samples. The optimum melting conditions were determined by variation of laser pulse duration, sample size, and thickness of the pressure-transmitting (NaCl) medium.

As sample for the diamond cell study, we chose Ge for several reasons. First, we thought Ge, which has a larger effective hard core than Si but the same sp^3 -based tetrahedral bonding, might be closer in character to the model monatomic glassformer of our earlier (zero-pressure) study⁵, as indeed we verify below. Second, Ge, in the normal-pressure (1 atm) liquid state, has a higher first-shell coordination number than liquid Si, and is considered to resemble a disordered β -tin structure¹⁵. It has a metallic value and temperature coefficient of electrical conductivity¹⁶, and should behave more obviously like a metal under high-pressure conditions where the electron delocalization must be higher still. Third, partial vitrification of Ge in belt-type anvil press quenches has been reported¹⁷. Finally, Ge has little affinity for carbon, and hence is less likely to chemically damage the DAC windows during melting experiments. Subsequently, the molecular dynamics of liquid Ge (discussed later) yields a diffusivity of only $0.76 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ at T_3 , even lower than for Si.

Using the sample mountings described in the Methods section, we prepared 20- μm -diameter Ge samples, melt-quenched from pressures in the range 5–11 GPa (the upper limit for our diamond culets), for *in situ* Raman, and *ex situ* transmission electron microscope (TEM) examination. The Raman spectra for pressures above 7.9 GPa (by ruby fluorescence) showed the disappearance of the sharp intense line at 326 cm^{-1} (298 cm^{-1} at 0 GPa) and its replacement by a featureless broad fluorescence band, from which no structural details could be obtained.

The pressures, after the quench, were found to differ from the initial value by $\pm 0.8 \text{ GPa}$, so we report results as the initial pressure with the final pressure in parentheses. In repeat experiments, glasses are always obtained at 7.9 GPa and above, ~11 GPa being our

¹Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, USA. ²Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850, USA. ³J. Nehru Centre for Advanced Scientific Research, Jakkur Campus, Bangalore 560064, India.

maximum. Four trials at ~ 5 GPa (including one using green laser heating) failed to vitrify. Attempts to characterize a vitreous sample *in situ* (under pressure) by X-ray diffraction using the Sector 1 Advanced Photon Source (APS) at Argonne, unfortunately were not successful owing to the very small sample size. The TEM results, by contrast, provided detailed evidence that vitreous Ge was produced directly from the melt in the case of quenches of small samples at pressures near T_3 .

Figure 1a and b shows twin optical microscope images of the DAC aperture with embedded Ge crystals before, and spread glass after, the melting pulse. The TEM image in Fig. 1c shows a ~ 300 -nm-tall

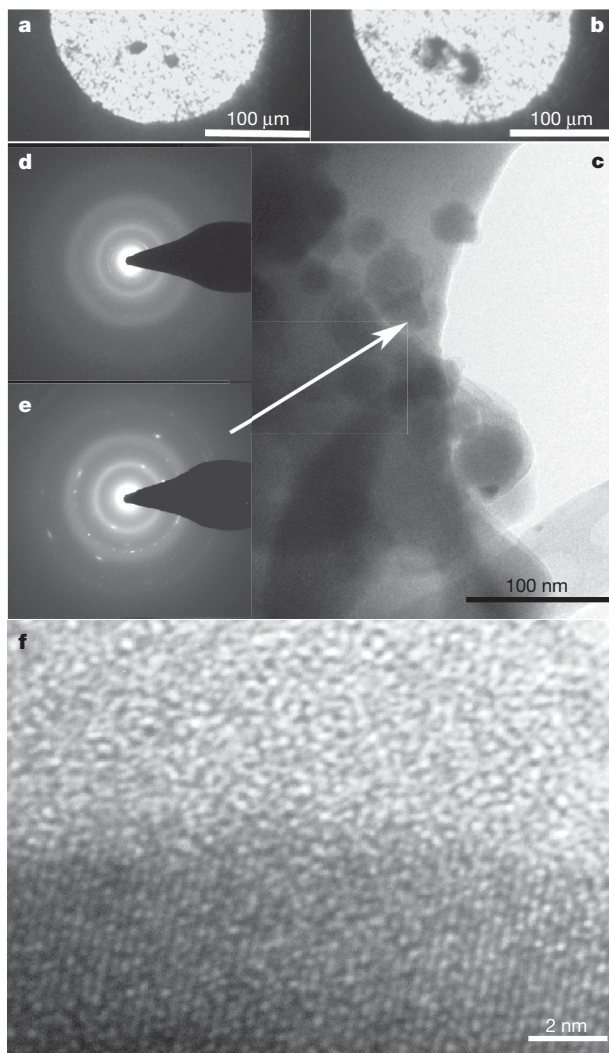


Figure 1 | Optical and electron micrographs of vitrified Ge. **a, b,** The Ge sample through the gasket aperture (between the diamond anvils) before and after the melting pulse. **c,** 300 nm segment of the 2 μ m sample fragment of vitrified Ge quenched at 7.9(7.1) GPa, showing globules in matrix. Energy-dispersive X-ray analysis of both shows only Ge and Cu (from the grid) to be present. **d,** Amorphous diffraction pattern obtained from all areas except the area including the globule indicated by the arrowhead. **e,** Amorphous diffraction pattern for the globule area. Laue spots, indicating crystalline character, are obvious. The crystal–glass boundary (at the arrow tip) is seen in the high-resolution image of **f.** The crystal has grown within a globule. Energy-dispersive X-ray analysis of both shows only Ge and Cu (from the grid) present (see Supplementary Information). Other globules in **c** are fully amorphous. This is reminiscent of the preferential crystallization of the LDA in Al_2O_3 – Y_2O_3 polyamorphic transitions²⁹, and also in the molecular triphenyl phosphite case³⁰. Annular dark-field imaging (see Supplementary Information) proves that the globules in **c** are darker because they are thicker, protruding from the matrix as seen at the right edge of **c.** In the annular dark field they appear bright.

section of a 2 μ m sample from a quench at 7.9(7.1) GPa, with two adjacent electron diffractograms (Fig. 1d and e). Figure 1d is from one of several electron-thin regions that gave the same diffraction pattern. Also, in subsequent studies of samples melt-quenched in the favourable pressure range, the same pattern is obtained. It is clearly an ‘amorphous’ pattern, being devoid of any irregularities or inhomogeneities that could be attributed to nanocrystals. Figure 1e is taken from a small area that includes a globule of interest to our discussion (at arrowhead), and contains a number of spots in addition to the diffuse rings. The spots are due to a single diamond-cubic Ge crystal originating within the globule (see high-resolution image in Fig. 1f). Globules with and without similar embedded crystals were seen in other sections of the 2 μ m fragment (Fig. 1c, which is a small part of the sample in Fig. 1b). The TEM image (Fig. 1c) was chosen over those from other successful runs (all of which gave the same amorphous diffractograms), because of the presence of the globular features—see our discussion below. The globular sample, the importance of which we discuss below, was obtained at the low-pressure limit (7.9(7.1) GPa) of the range for successful vitrifications. All other vitrified samples had uniform textures. The structure factor for the glassy phase has been derived from the diffractogram and is compared with other relevant data in Fig. 2.

Thus far we have tested the vitrifiability of liquid Ge near its triple point. Clearly metallic, even at normal pressure, liquid Ge yields vitreous states when melt-quenched in the DAC from near and somewhat above the triple point pressure, but not from much below it. The possibility that the quench product is a β -tin-like (or other high-density) crystal, which then amorphizes during decompression is eliminated by our temperature of decompression, 25 $^\circ\text{C}$, because much work has shown that ambient decompression of metastable metal crystals of this type, including Ge¹⁸, yields crystalline semiconductors. Details on five such possible paths for Ge, all documented, all with the same conclusion, are given in Supplementary

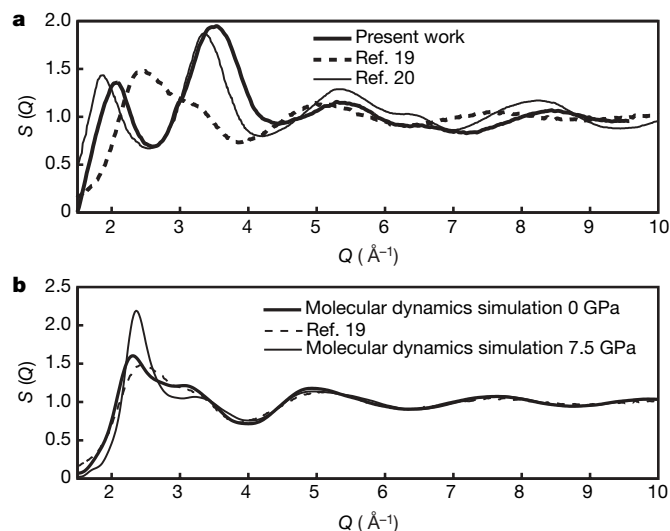


Figure 2 | Comparisons of structure factors for vitreous and liquid states of Ge from laboratory and molecular dynamics simulation studies.

a, Structure factor $S(Q)$ (that is, the scattered intensity at the scattering vector Q) for glassy Ge derived from the electron diffraction pattern of Fig. 1 (thick solid line) compared with the corresponding TEM electron-diffraction-based $S(Q)$ for vapour-deposited amorphous Ge²⁰ (thin solid line). The thick dashed line is $S(Q)$ at 1,273 K from liquid Ge¹⁹. The similarity of the present $S(Q)$ to that of the vapour-deposited value confirms the LDA character, while the displacement of the peaks to higher Q for our sample reflect its formation at pressure exceeding zero. **b,** Our Ge simulation results for the liquid at zero pressure (thick solid line), and at 7.5 GPa (thin solid line) showing agreement with laboratory results¹⁹ at zero pressure and 1,273 K (dashed line). Small Q changes reflect mainly changes in second-, third- and higher-shell coordination numbers.

Information. We conclude that our successful quenches have yielded the first examples of vitrification of a monatomic metallic liquid.

However, just as the high-pressure crystals are known not to survive decompression to ambient pressure unless first cooled below 100 K, so also a metallic glass does not survive our ambient-temperature decompression. Rather, it transforms from a high-density amorphous (HDA) glass to a low-density amorphous (LDA) glass. We now verify that we have observed the LDA (in Fig. 1). To aid our analysis we use additional simulations on a model of Ge (V.M., manuscript in preparation) similar to one used to study Ge surfaces (see Supplementary Information) for which no comparisons with experiments were made. We make comparisons in Fig. 2.

Figure 2b shows the excellent agreement of calculated and observed¹⁹ structure factors of liquid Ge at normal pressure, and also the effect of pressure increase to 7.5 GPa. The latter effect is confirmed by *ab initio* molecular dynamics studies¹⁵ analysed in terms of increasing metallicity. A glass formed from this liquid should be unambiguously metallic. However, the Fig. 2a structure factor for our glass, derived from the Fig. 1d diffraction pattern, is strikingly different from that of the liquid¹⁹. As anticipated above, it closely resembles that obtained from electron diffraction of vapour-deposited Ge²⁰ (the LDA non-metallic form), and also from X-ray scattering for vapour-deposited²¹ and electrodeposited Ge (see Supplementary Information). The differences between the high- and low-density forms are shared by water, Si and Ge²².

The identification of our glass with the previous LDAs seems unambiguous. Our finding is consistent with that of ref. 13 in the only parallel of our work—the high-pressure vitrification of liquid GaSb. (GaSb at 1 atm is a zincblende phase with a low melting point of 970 K, decreasing with increasing pressure to 670 K at T_3). Again, the dense polymorph does not survive decompression at room temperature. Indeed, this is expected from the theoretical model²³ which first predicted the liquid–liquid transitions: Aptekar²³ showed a spinodal limit on the stability of both HDA Ge and Si at 2–4 GPa (see also Supplementary Information).

To identify the liquid–liquid transition, we turn to a composite phase diagram (Fig. 3) that combines what we know from the earlier simulation study⁵ with what we know from laboratory high-pressure²⁴ and supercooling studies²⁵. Figure 3 shows a temperature–pressure phase diagram for Ge, whose vertical axis is projected from the temperature– λ phase diagram⁵ at $\lambda = 20$ (versus $\lambda = 21$ for Si), as shown to be appropriate by the $\langle S(Q) \rangle$ comparisons in Fig. 2b. (λ is the ‘tetrahedrality parameter’ in the Stillinger–Weber potential^{5,11}.) We emphasize the projection of the liquid–liquid transition line of the temperature– λ phase diagram into the pressure plane of the temperature–pressure diagram.

Pressure increases depress not only the melting point of Ge, but also the liquid–liquid transition²⁶, as predicted in refs 23 and 27 and as observed by molecular dynamics for Stillinger–Weber Si (ref. 12; V.M., S.S. and C.A.A., unpublished work). We represent the glass transition temperature T_g by a thick square-dotted line, using the observation¹² (for Stillinger–Weber Si) that isothermal diffusivity passes through a maximum below the T_3 pressure and that T_g must be above 473 K at 8.1(8.9) GPa, because heating an *in situ* vitrified sample up to 473 K (the pressurized cell safety limit) for one hour did not cause crystallization (according to Raman spectra). We represent the quenches of this study on the phase diagram of Fig. 3 (see lines a, b and c).

The lower-pressure, crystallizing, melt-quenches of our current experiments are represented by the vertical dashed line ‘a’ that passes through the liquid–liquid line while the liquid is still well above the extension of the T_g line. Crystals form rapidly as they cross the liquid–liquid line^{5,12,26} because the gap in configuration space between LDA and crystal is so much smaller than that between HDA and the crystal¹⁵.

The melt-quenches that produce amorphous Ge without globules are represented by vertical line ‘c’. Above T_3 , the liquid–liquid line is almost vertical (see caption and Supplementary Information). Thus, above T_3 there is no possibility that the liquid can transform to LDA before it vitrifies. The homogeneous glass obtained at 10.6(10.1) GPa

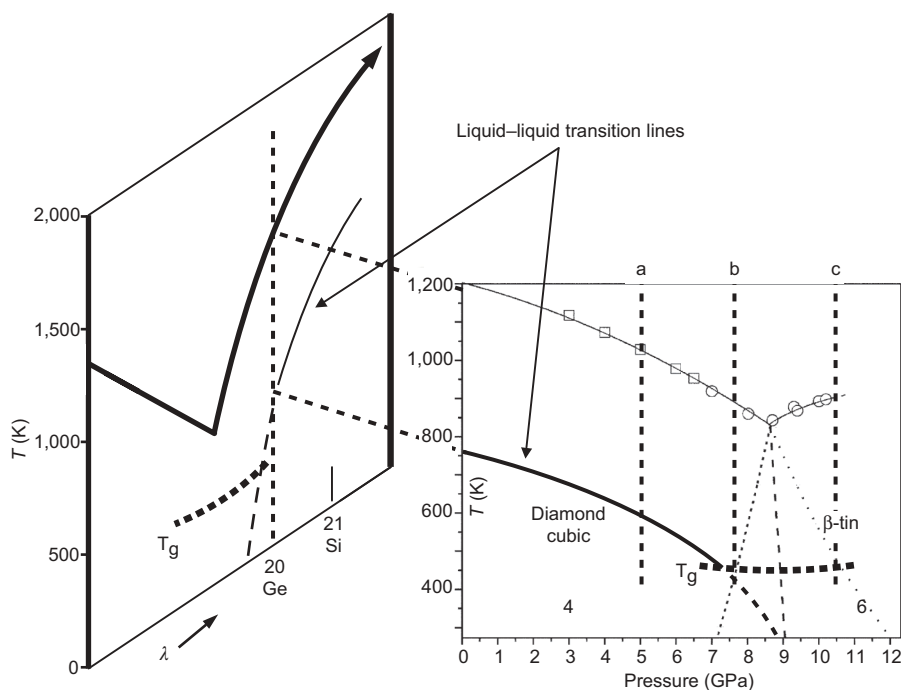


Figure 3 | Relation of temperature–pressure phase diagram for Ge to the temperature– λ potential diagram for Stillinger–Weber systems. The temperature–pressure phase diagram for Ge²⁴ is shown in relation to the temperature– λ phase diagram of ref. 5 by which the relation of Ge to Si in the temperature– λ diagram is understood. The temperature scale changes for Ge, and becomes that of the laboratory Ge phase diagram, and the gap between T_m and the liquid–liquid transition temperature T_{LL} has been set to the experimental supercooling limit found by cooling small levitated samples²⁵. Projections of the temperature– λ diagram at $\lambda = 20$ into the pressure plane show how the T_m and T_{LL} transition lines of Ge change with pressure. The T_{LL} line is shown approaching the vertical at the T_3 pressure, for thermodynamic reasons given in ref. 5. In water polymorphism³², T_{LL} , which goes vertical at the T_3 pressure, tracks the homogeneous crystal nucleation line T_h . T_h , much studied in solution systems where composition replaces the pressure axis, goes almost vertical near the eutectic composition (the equivalent of the T_3 pressure), intersecting the ideal glass and Kauzmann temperatures T_0 and T_K at the eutectic temperature (see Supplementary Information). Therefore, LDA cannot form during cooling above the T_3 pressure, for example, along line ‘c’. For a description of the vertical dashed lines ‘a’, ‘b’ and ‘c’, see text. The numbers ‘4’ and ‘6’ indicate the coordination numbers of crystal phases.

must therefore be the HDA metallic form. Thus we have vitrified a metallic liquid into a metallic glass. As argued above (see details in Supplementary Information), this metallic glass transforms to LDA during ambient decompression, via spinodal collapse^{23,28}.

The vertical line 'b' is the cooling trajectory of the sample that yielded the globule phase as a consequence of passing the liquid–liquid line and the T_g line almost simultaneously. Then, as seen in Y_2O_3 – Al_2O_3 glass²⁹ and also in triphenyl phosphite³⁰, nucleated droplets of the LDA phase are vitrified as they form (LDA being less diffusive than HDA). Given the propensity of low-temperature liquid polyamorphs to crystallize^{29,30}, the existence of nanocrystals within many, but not all, of the amorphous globules then provides circumstantial microscopic evidence for the controversial liquid–liquid transition in Si-like supercooled liquids^{12,26,31}.

METHODS SUMMARY

DAC studies. Crystalline Ge samples were embedded in a thin layer of pressure-distributing NaCl inside a 210 μ m hole in a T-301 stainless steel gasket, which was mounted in a Merrill-Bassett DAC. Ruby chips were included for pressure determination using the fluorescence spectrum frequency shift. A Synrad 60-1 CO_2 laser provided 125 W pulses of duration 2–100 μ s in trial experiments. Glass formation was initially inferred from the disappearance of the strong Raman signal for diamond-cubic Ge at 326 cm^{-1} (7.9 GPa). The Ge sample (Fig. 1b), embedded in NaCl, was then carefully transferred onto a TEM copper grid (Ted Pella, Inc.) with lacey carbon support, and the NaCl carefully dissolved away using distilled water, which usually leaves the Ge on the grid. The high-resolution images and selected area diffraction patterns were recorded with a Philips CM200 field emission gun electron microscope operated at 200 kV. The instrument is equipped with a Gatan slow-scan charge-coupled device (CCD) camera, an energy-dispersive X-ray detector (EDAX) and the Gatan-Digital Micrograph and ES Vision acquisition systems.

Data treatment. Diffraction data for amorphous Ge were collected on the above-mentioned high-resolution TEM and calibrated using an Au standard. The two-dimensional diffraction image was radially integrated and scaled to the Dirac–Fock atomic form factors (200 keV electrons)³³. The resulting structure factor $S(Q)$ is shown in Fig. 2.

Molecular dynamics simulations summary. The Stillinger–Weber Si simulations¹¹ were performed with protocols described in ref. 4. For Ge, periodic isobaric-isothermal simulations of 5,000 atoms modelled with the Stillinger–Weber¹¹ potential modified with $\lambda = 20$, $\epsilon = 45$ kcal mol^{−1} and $\sigma = 2.1836$ Å were carried out using the LAMMPS code (see Methods and Supplementary Information).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 February; accepted 21 June 2007.

1. Privalko, Y. Excess entropies and related quantities in glass-forming liquids. *J. Phys. Chem.* **84**, 3307–3312 (1980).
2. Martinez, L. M. & Angell, C. A. A thermodynamic connection to the fragility of glass-forming liquids. *Nature* **410**, 663–667 (2001).
3. Johnson, W. L. Bulk glass-forming metallic alloys: science and technology. *Mater. Res. Soc. Bull.* **24**, 42–50 (1999).
4. Wang, W. H., Lewandowski, J. J. & Greer, A. L. Understanding the glass-forming ability of Cu₅₀Zr₅₀ alloys in terms of a metastable eutectic. *J. Mater. Res.* **20**, 2307–2313 (2005).
5. Molinero, V., Sastry, S. & Angell, C. A. Tuning of tetrahedrality in a silicon potential yields a series of monatomic (metal-like) glass-formers of very high fragility. *Phys. Rev. Lett.* **97**, 075701 (2006).
6. Cohen, M. H. & Turnbull, D. Composition requirements for glass formation in metallic and ionic systems. *Nature* **189**, 131–132 (1961).
7. van Ketel, W., Das, C. & Frenkel, D. Structural arrest in an ideal gas. *Phys. Rev. Lett.* **94**, 135703 (2005).
8. Hilsch, R. in *Non-Crystalline Solids* (ed. Frechette, V. D.) 348 (J. Wiley and Sons, New York, 1960).
9. Chathoth, S. M., Meyer, A., Koza, M. M. & Juranyi, F. Atomic diffusion in liquid Ni, NiP, PdNiP, and PdNiCuP alloys. *Appl. Phys. Lett.* **85**, 4881–4883 (2004).
10. Wachtel, E. et al. Magnetic-susceptibility and DSC study of the crystallization of melt-quenched Ni–P amorphous-alloys. *Mater. Sci. Eng. A* **133**, 196–199 (1991).
11. Stillinger, F. H. & Weber, T. A. Computer-simulation of local order in condensed phases of silicon. *Phys. Rev. B* **31**, 5262–5271 (1985).

12. Angell, C. A., Borick, S. & Grabow, M. Glass transitions and first order liquid-metal-to-semiconductor transitions in 4–5–6 covalent systems. *J. Non-Cryst. Solids* **207**, 463–471 (1996).
13. Brazhkin, V. V., Larchev, V. I., Popova, S. V. & Skrotskaya, G. G. The influence of high pressure on the disordering of the crystal structure of solids rapidly quenched from the melt. *Phys. Scr.* **39**, 338–340 (1989).
14. Tanaka, H. Simple view of waterlike anomalies of amorphous liquids with directional bonding. *Phys. Rev. B* **66**, 064202 (2002).
15. Koga, J., Nishio, K., Yamaguchi, T. & Yonezawa, F. Tight-binding molecular dynamics study on the structural change of amorphous germanium with the increase of density. *J. Phys. Soc. Jpn* **73**, 388–396 (2004).
16. Schnydersy, H. S. & Van Zytveldt, J. B. Electrical resistivity and thermopower of liquid Ge and Si. *J. Phys. Condens. Matter* **8**, 10875–10883 (1996).
17. Zhang, F. X. & Wang, W. K. Microstructure of germanium quenched from the undercooled melt at high-pressures. *Appl. Phys. Lett.* **67**, 617–619 (1995).
18. Brazhkin, V. V., Lyapin, A. G., Popova, S. V. & Voloshin, R. N. Non-equilibrium phase transitions and amorphization in Si, Si/GaAs, Ge, and Ge/GaSb at the decompression of high-pressure phases. *Phys. Rev. B* **51**, 7549–7554 (1995).
19. Salmon, P. S. A. Neutron-diffraction study on the structure of liquid germanium. *J. Phys. F* **18**, 2345–2352 (1988).
20. Ankele, J., Mayer, J., Lamparter, P. & Steeb, S. Structure factor of amorphous-germanium by quantitative electron-diffraction. *J. Non-Cryst. Solids* **193**, 679–682 (1995).
21. Etherington, G. et al. A neutron-diffraction study of the structure of evaporated amorphous-germanium. *J. Non-Cryst. Solids* **48**, 265–289 (1982).
22. Benmore, C. J. et al. Intermediate range chemical ordering in amorphous and liquid water, Si, and Ge. *Phys. Rev. B* **72**, 132201(4) (2005).
23. Aptekar, L. I. Phase transitions in non-crystalline germanium and silicon. *Sov. Phys. Dokl.* **24**, 993–995 (1979).
24. Voronin, G. A. et al. In situ X-ray diffraction study of germanium at pressures up to 11 GPa and temperatures up to 950 K. *J. Phys. Chem. Solids* **64**, 2113–2119 (2003).
25. Li, D. & Herbach, D. M. Containerless solidification of germanium by electromagnetic levitations and in a drop-tube. *J. Mater. Sci.* **32**, 1437–1442 (1997).
26. Sastry, S. & Angell, C. A. Liquid–liquid phase transition in supercooled liquid silicon. *Nature Mater.* **2**, 739–743 (2003).
27. Ponyatovsky, G. G. A thermodynamic approach to T–P phase diagrams of substances in liquid and amorphous states. *J. Phys. Condens. Matter* **15**, 6123–6141 (2003).
28. Mishima, O. Reversible transition between two H₂O amorphs at ~0.2 GPa and ~135 K. *J. Chem. Phys.* **100**, 5910–5919 (1991).
29. Aasland, S. & McMillan, P. F. Density-driven liquid–liquid phase-separation in the system Al₂O₃–Y₂O₃. *Nature* **369**, 633–636 (1994).
30. Kurita, R. & Tanaka, H. Critical-like phenomena associated with liquid–liquid transition in a molecular liquid. *Science* **306**, 845–848 (2004).
31. Spaepen, F. & Turnbull, D. in *Laser-Solid Interactions and Laser Processing 1978* (eds Ferris, S. D., Leamy, H. J. & Poate, J.) AIP Conf. Proc. **50**, 73–83 (1979).
32. Stanley, H. E. et al. The puzzling behavior of water at very low temperature. *Phys. Chem. Chem. Phys.* **2**, 1551–1558 (2000).
33. Rez, D., Rez, P. & Grant, I. Dirac–Fock calculations of X-ray scattering factors and contributions to the mean inner potential for electron scattering. *Acta Crystallogr. A* **50**, 481–497 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Benmore and Q. Mei for the attempted *in situ* studies of melt-quenched samples at the Argonne APS. We acknowledge the allocation of computer time at the Center for High Performance Computing at the University of Utah, where the simulations on Ge were carried out. The cooperation of D. Matyushov and his group (whose Arizona State University computing facilities were used for the Si simulations) is also appreciated. We also thank our NSF-CRC colleagues P. Debenedetti, G. Stanley and P. Rossky for discussions. This work was supported by NSF grants from the Chemistry CRC (to C.A.A.), the DMR Solid State Chemistry (to C.A.A.), the NSF Chemistry (to J.L.Y.), the Carnegie/DOE Alliance Center (DOE–NNSA CDAC) (to J.L.Y.) and the Swarnajayanti Fellowship, DST, India (to S.S.). The TEM studies depended on the John M. Cowley Center for High Resolution Electron Microscopy.

Author Contributions C.A.A., V.M. and S.S. conceived the project as part of previous work⁵. J.L.Y. recommended and directed the DAC investigation, M.H.B. and E.S. executed the sample mounting, laser pulse melting, and Raman characterization experiments, V.C.S. and M.H.B. performed the TEM studies, V.M. carried out both preliminary Si and later Ge molecular dynamics simulations, and wrote the results analysis given in the Supplementary Information, and C.A.A. wrote the paper (with advice and criticism from all co-authors).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.A.A. (caa@asu.edu).

METHODS

Molecular dynamics simulations on Stillinger–Weber Si and Ge. Molecular Dynamics (MD) simulations of Si with the Stillinger–Weber¹¹ potential were executed at different pressures from 0 to 16 GPa, and those carried out in the vicinity of the triple point (~ 11 GPa) and up to 16 GPa never crystallized. Higher pressures were not investigated. The diffusivity at the triple point was $0.96 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$.

The simulations presented in the letter were carried on a model of Ge based on the Stillinger–Weber potential for Si with tetrahedrality parameter $\lambda = 20$, and scaling the two-body attraction term to reproduce the crystal energy of Ge, $\varepsilon = 45 \text{ kcal mol}^{-1}$ and $\sigma = 2.1836 \text{ \AA}$ (V.M., manuscript in preparation). These runs were executed using a fast parallel-processing MD code (LAMMPS code³⁴), and periodic boundary conditions, with isobaric–isothermal simulations on 5,000 atoms. A similar model with $\lambda = 19.5$ has been used to study solid Ge surfaces³⁵, and was recently applied to liquid Ge–Si alloys³⁶.

As seen in Fig. 2b, the agreement of the Ge simulations with experiment is very good, minor differences only showing up at small Q , where the second and third and higher neighbour shells are involved. According to this model, the diffusivity at the triple point, $0.76 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$, is even lower than it is for Si, implying that crystallization would never be observed on any computational timescale.

34. Plimpton, S. Fast parallel algorithms for short-range molecular-dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
35. Grabow, M. H. & Gilmer, G. H. Thin-film growth modes, wetting and cluster nucleation. *Surf. Sci.* **194**, 333–346 (1988).
36. Yu, W. B. & Stroud, D. Molecular-dynamics study of surface segregation in liquid semiconductor alloys. *Phys. Rev. B* **56**, 12243–12249 (1997).

Indirect radiative forcing of climate change through ozone effects on the land-carbon sink

S. Sitch¹, P. M. Cox³, W. J. Collins⁴ & C. Huntingford²

The evolution of the Earth's climate over the twenty-first century depends on the rate at which anthropogenic carbon dioxide emissions are removed from the atmosphere by the ocean and land carbon cycles¹. Coupled climate–carbon cycle models suggest that global warming will act to limit the land-carbon sink², but these first generation models neglected the impacts of changing atmospheric chemistry. Emissions associated with fossil fuel and biomass burning have acted to approximately double the global mean tropospheric ozone concentration³, and further increases are expected over the twenty-first century⁴. Tropospheric ozone is known to damage plants, reducing plant primary productivity and crop yields⁵, yet increasing atmospheric carbon dioxide concentrations are thought to stimulate plant primary productivity⁶. Increased carbon dioxide and ozone levels can both lead to stomatal closure, which reduces the uptake of either gas, and in turn limits the damaging effect of ozone and the carbon dioxide fertilization of photosynthesis⁶. Here we estimate the impact of projected changes in ozone levels on the land-carbon sink, using a global land carbon cycle model modified to include the effect of ozone deposition on photosynthesis and to account for interactions between ozone and carbon dioxide through stomatal closure⁷. For a range of sensitivity parameters based on manipulative field experiments, we find a significant suppression of the global land-carbon sink as increases in ozone concentrations affect plant productivity. In consequence, more carbon dioxide accumulates in the atmosphere. We suggest that the resulting indirect radiative forcing by ozone effects on plants could contribute more to global warming than the direct radiative forcing due to tropospheric ozone increases.

Tropospheric ozone (O_3) is a naturally occurring greenhouse gas formed as a product of photochemical reactions with precursors: NO_x , CH_4 , CO and volatile organic compounds. Over the industrial period, anthropogenic precursor emissions from fossil fuel and biomass burning have led to elevated ambient ozone concentrations ($[O_3]$) over a large portion of the Earth's surface (Fig. 1), resulting in a direct radiative forcing of climate change of $+0.35^{+0.1}_{-0.3} \text{ W m}^{-2}$ (ref. 8). Many regions of the globe are already experiencing near-surface ozone levels greater than 40 parts per billion (40 p.p.b.)—levels that may cause visible leaf injury and plant damage, and reduction in crop yields⁵, with associated economic costs of several billion dollars per annum in each of the US, EU and East Asia^{5,9}.

Concentrations of atmospheric CO_2 and near surface ozone are expected to increase significantly through the next century, although the magnitude of the increases depends on the particular emission scenario chosen. Here we use the SRES A2 scenario for comparability with recent atmospheric chemistry simulations¹⁰, but our overall conclusion is insensitive to this choice. Under SRES A2, mean monthly 24-h ozone concentrations by 2100 are projected to be

above 40 p.p.b. over almost all regions, and to exceed 70 p.p.b. over western and central Eurasia, eastern and western North America, Brazil, central and southwestern Africa, and East Asia, during the Northern Hemisphere summer (Fig. 1).

Ozone causes cellular damage inside leaves that adversely affects plant production, reduces photosynthetic rates and requires increased resource allocation to detoxify and repair leaves⁵. There have been few global modelling studies of the impact of tropospheric ozone on plant production and global land-carbon storage¹¹, and no study has estimated the indirect radiative forcing of tropospheric ozone through feedbacks on the global carbon cycle. Here, we are concerned with the possible impacts of future tropospheric ozone on global-scale plant primary production, land-carbon storage, and its implications for twenty-first-century climate change.

Future elevated $[CO_2]$ may itself lead to reductions in stomatal conductance¹² at levels that act to alleviate future O_3 plant damage. Hence, future O_3 effects on plants are defined by the interplay of ambient $[O_3]$, $[CO_2]$ and climate change on stomatal conductance and plant production, with important ramifications for global land-carbon and hydrological cycles^{7,11}. Free air CO_2 enrichment (FACE) experiments and other ambient air experiments indicate a nonlinear interaction between plant responses to CO_2 and O_3 (refs 6, 13–15). To account for these interactions, we use a flux-gradient approach to modelling ozone damage¹⁶, rather than the more usual empirical approach based on the accumulated ozone exposure above 40 p.p.b. (ref. 11). We modify the MOSES-TRIFFID land-surface scheme¹⁷, assuming a suppression of net leaf photosynthesis by ozone that varies proportionally to the ozone flux through stomata above a specified critical ozone deposition flux. Our scheme includes an empirical relationship between stomatal conductance and photosynthesis¹⁸, and through this mechanism the direct effect of O_3 deposition on photosynthesis also leads to a reduction in stomatal conductance. As the O_3 flux itself depends on the stomatal conductance, which in turn depends upon the net rate of photosynthesis¹⁸, the model requires a consistent solution for the net photosynthesis, stomatal conductance and the ozone deposition flux (see Methods).

This more mechanistic approach to modelling ozone effects on photosynthesis accounts for the complex interaction between CO_2 and O_3 effects. Increases in the concentration of either gas lead to stomatal closure, which may limit the uptake of the other gas. Thus the model can be used to assess the extent to which CO_2 -induced stomatal closure⁷ will protect plants against the potentially damaging impacts of increases in near-surface O_3 , or conversely the extent to which O_3 increases will limit CO_2 -fertilization of photosynthesis and thereby reduce the ability of ecosystems to mitigate global warming. In this study, we focus on the interplay between future projections of O_3 and CO_2 on plant physiology and the land carbon cycle, but acknowledge the importance of the additional and uncertain

¹Met Office, Hadley Centre for Climate Prediction and Research (JCHMR), Met Office Building, Met Office, Exeter, EX1 1PB, UK. ²Centre for Ecology and Hydrology Wallingford, Wallingford, OX10 8BB, UK. ³School of Engineering, Computer Science and Mathematics, University of Exeter, Exeter, EX4 4QF, UK. ⁴Met Office, Hadley Centre for Climate Prediction and Research, Fitzroy Road, Exeter, EX1 3PB, UK.

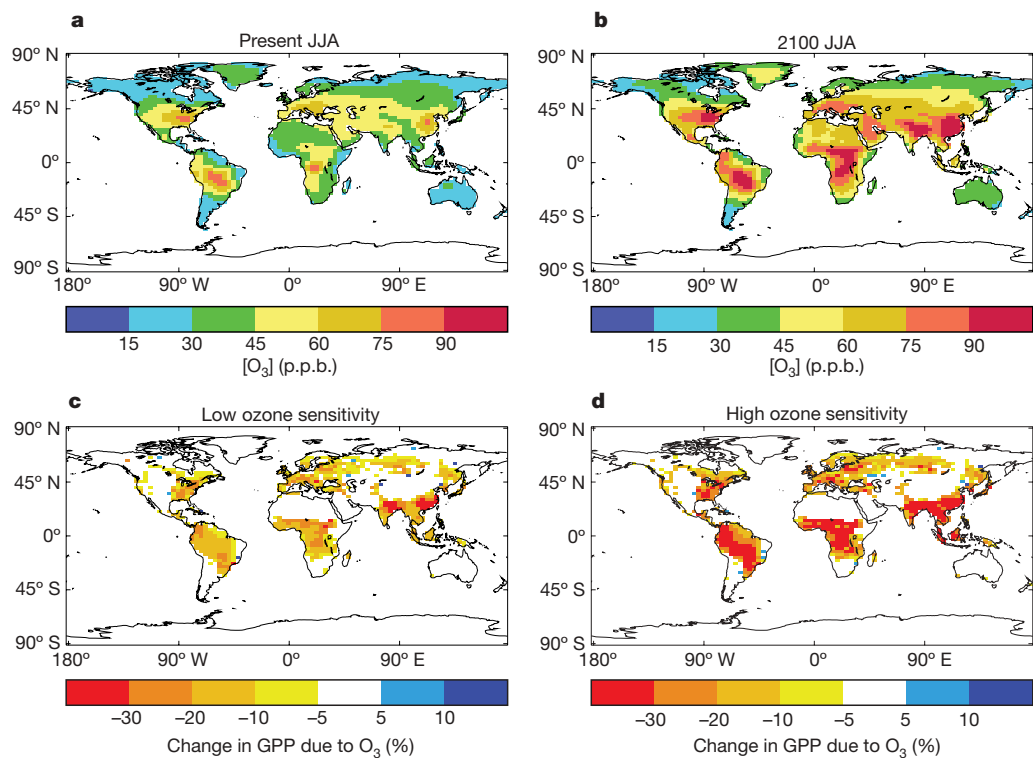


Figure 1 | Temporal changes of modelled ozone concentrations and gross primary productivity. **a, b,** Modelled diurnal (24-h) mean surface [O₃] in p.p.b. averaged over June, July and August (JJA) for the present day (**a**) and the year 2100 under the SRES A2 emissions scenario (**b**). **c, d,** Simulated

percentage change in gross primary productivity (GPP) between 1901 and 2100 due to O₃ effects at fixed pre-industrial atmospheric [CO₂] for ‘low’ (**c**) and ‘high’ (**d**) ozone plant sensitivity.

interactions between a changing climate, stomatal conductance, ozone uptake and plant productivity.

We use spatially explicit [O₃] fields derived from the STOCHEM model¹⁹ to drive the modified MOSES-TRIFFID land-surface scheme¹⁷ offline. Two model versions are applied, with ‘high’ and ‘low’ plant ozone sensitivity based on observations^{16,20}. For model evaluation against global carbon cycle budgets, simulations are conducted for the period 1901–2002, using a monthly observational data set of the twentieth-century climate²¹, changing monthly fields of diurnal mean [O₃] and prescribed annual fields of global atmospheric [CO₂]. As a further model evaluation, a second set of simulations replicates the Aspen FACE experiment^{6,13} and the free-air O₃ fumigation experiment at a Swiss grassland site¹⁴. The Aspen FACE experiment investigates the response of maturing aspen stands across a wide range of O₃ sensitivity in five aspen clones. Plots were planted in 1997 and exposed for seven years to combinations of ambient and elevated concentrations of [CO₂] and [O₃], with elevated [CO₂] at 560 p.p.m.v., and elevated [O₃] at 1.5 times ambient levels. In the grassland experiments, plots are exposed for 5 years to ambient and elevated [O₃] at 1.5 times ambient levels. Our simulations broadly agree with the results from these free air enrichment Swiss grassland and aspen experiments^{6,13,14} (see Supplementary Fig. 3), and an ozone risk mapping for North America derived from the Aspen FACE results¹⁵.

A third set of factorial simulations is conducted from 1901 to 2100 with changing fields of monthly near-surface tropospheric ozone and atmospheric CO₂ concentrations consistent with the A2 SRES scenario. A fixed mean monthly pre-industrial climate is prescribed in these runs, in order to focus on the interaction between direct CO₂ and O₃ effects on plant physiology. For each of the ‘high’ and ‘low’ ozone plant sensitivity parameterizations, three simulations are conducted with combinations of fixed pre-industrial and prescribed changing fields of [O₃] and global atmospheric [CO₂].

Over the 1990s, global mean land-atmosphere fluxes of $-1.34 \text{ Pg C yr}^{-1}$ and $-1.74 \text{ Pg C yr}^{-1}$ are simulated for the ‘high’

and ‘low’ plant O₃ sensitivity models, respectively, both within the IPCC range²² of $-4.3 \text{ Pg C yr}^{-1}$ to $-0.9 \text{ Pg C yr}^{-1}$ with a mean of $-2.6 \text{ Pg C yr}^{-1}$ (Supplementary Table 2). Figure 1 shows the impact of O₃ increases on the pattern of gross primary productivity (GPP) by 2100. The lower panels show the percentage change in GPP due to O₃ in the model runs with fixed pre-industrial CO₂. Over the period 1901–2100, global GPP is projected to decrease by 14–23% owing to plant ozone damage (Supplementary Table 3), with regional reductions above 30% (Fig. 1). Large reductions in GPP and land-carbon storage are projected over North America, Europe, China and India, regions with the highest levels of human appropriation of primary productivity²³, and in tropical ecosystems, raising important issues concerning the vulnerability of regional ecosystem services (for example, food security, forest productivity and carbon sequestration)²⁴ to changes in air quality.

The combined effect of elevated future [CO₂] and [O₃] on plant physiology is an increase in global GPP and net land carbon uptake (Table 1). However, the enhancement in global GPP at 2100, by the

Table 1 | Simulated changes in the global land carbon cycle due to O₃ and CO₂ increases

Model	GPP (Pg C yr ⁻¹)	Veg. C (Pg C)	Soil C (Pg C)	Land C (Pg C)
‘High’ plant O₃ sensitivity				
Value in 1901	112.7	461.8	1,110.8	1,572.6
Δ Value (2100–1901)				
Δ[CO ₂], fixed [O ₃]	88.4	235.0	621.7	856.7
Fixed [CO ₂], Δ[O ₃]	–26.4	–89.1	–173.4	–262.5
Δ[CO ₂] and Δ[O ₃]	58.4	184.8	432.7	617.5
‘Low’ plant O₃ sensitivity				
Value in 1901	116.6	488.2	1,130.2	1,618.4
Δ Value (2100–1901)				
Δ[CO ₂], fixed [O ₃]	86.9	217.5	618.3	835.8
Fixed [CO ₂], Δ[O ₃]	–16.1	–31.9	–111.3	–143.2
Δ[CO ₂] and Δ[O ₃]	71.3	201.9	513.8	715.6

Shown are changes (Δ) in global gross primary production (GPP) and global carbon stocks in vegetation (Veg.) and soils between 1901 and 2100.

physiological effects of elevated $[\text{CO}_2]$, is reduced by between $15.6 \text{ Pg C yr}^{-1}$ and $30.0 \text{ Pg C yr}^{-1}$, for the 'low' and 'high' plant ozone sensitivity, respectively, when concurrent ozone impacts on vegetation are also considered. This process is not accounted for in the current generation of climate-carbon cycle models².

A novel aspect of adopting a flux-gradient approach to modelling leaf gas exchange of both O_3 and CO_2 is that it enables investigation not only of the individual effects of O_3 and CO_2 acting in isolation on plant physiology, but also of their interaction.

Elevated CO_2 provides some protection against O_3 damage, by way of reductions in stomatal conductance and a concomitant decrease in plant ozone uptake. We diagnose this protection by comparing the O_3 -induced percentage reductions in GPP at 2100 from the runs with and without prescribed CO_2 increases (Supplementary Fig. 1, right-hand panels). CO_2 fertilization is projected to increase GPP significantly, so while the absolute effect of O_3 on GPP is greater under increasing rather than constant CO_2 , the fractional change is much smaller. CO_2 -induced stomatal closure is found to offset O_3 -suppression of GPP by over one-third, such that GPP by 2100 is 8–15% lower due to O_3 exposure, rather than 14–23% lower in the absence of CO_2 increases (see Supplementary Table 3). This

nonlinear interaction between the relative effects of CO_2 and O_3 increases is typically ignored in modelling approaches based on the cumulative O_3 exposure rather than uptake by leaves.

There are large potential impacts of elevated future $[\text{O}_3]$ on the ability of many ecosystems to sequester carbon (Table 1). Over the period 1900–2100, changes in $[\text{O}_3]$ with all other forcings fixed are projected to reduce land-carbon storage accumulation by between 143 Pg C and 263 Pg C (Fig. 2a, Supplementary Fig. 4). This is equivalent to a reduction of between 17% and 31% in the projected land uptake associated with the plant physiological response to CO_2 increase alone. However, there are significant uncertainties in the response of different plant species to O_3 , especially for tropical ecosystems, and in the scaling up of open top chamber results²⁵ (for example, uncertainties associated with chamber effects), both warranting further research. In our present study, ozone response data for European and North American species have been extrapolated to represent all global vegetation types.

Suppression of the land-carbon sink results in additional anthropogenic CO_2 emissions accumulating in the atmosphere, and therefore an indirect radiative forcing of climate change by O_3 effects on the terrestrial biosphere. The indirect radiative forcing due to O_3 is diagnosed for comparison to the direct forcing due to tropospheric O_3 (Fig. 2b), assuming 50% of the extra CO_2 from the land is sequestered by the oceans²⁶. The indirect forcing by 2100 is estimated at 0.62 W m^{-2} and 1.09 W m^{-2} for the 'low' and 'high' plant ozone sensitivity runs, respectively, which compares with a mean direct radiative forcing from 11 atmospheric chemistry models of 0.89 W m^{-2} (refs 3, 4, 27, 28). Although the absolute values of radiative forcing are dependent on our choice of emissions scenario, the relative importance of direct and indirect radiative forcing is much less sensitive to the uncertainty in emissions. As such, these results suggest that ozone effects on vegetation could double the effective radiative forcing due to increases in tropospheric ozone, significantly increasing the importance of changes in atmospheric chemistry as a driver of twenty-first-century climate change.

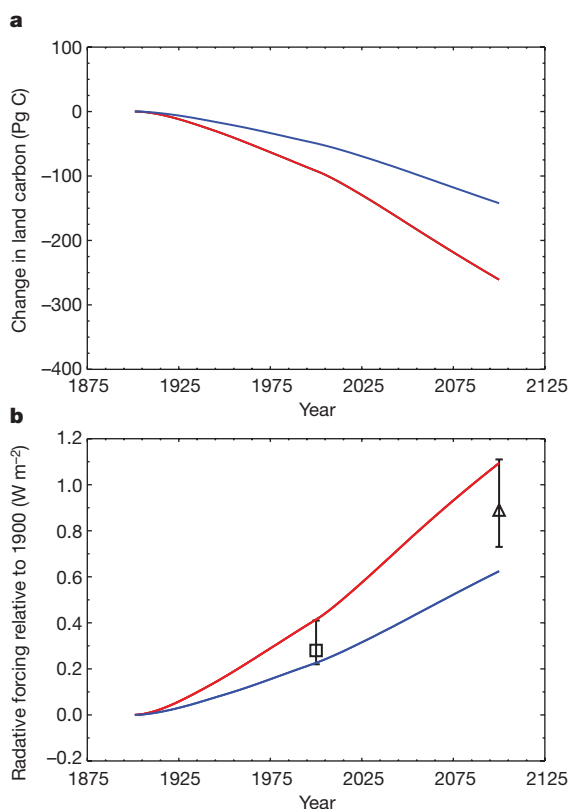


Figure 2 | Temporal changes in land carbon storage and radiative forcing due to ozone. **a**, **b**, Simulated change in land carbon storage (**a**) and indirect radiative forcing due to O_3 increases alone (**b**), for 'high' (red) and 'low' (blue) plant sensitivities to ozone. These results are diagnosed from model simulations using a fixed pre-industrial CO_2 concentration. For comparison, estimates of the direct radiative forcing due to O_3 increases are shown by the bars in **b**. Present-day direct radiative forcing comes from the STOCHEM-HadGEM1 tropospheric ozone fields (black square), with the bars showing the range of estimates from other atmospheric chemistry models³. Estimates of future direct radiative forcing due to O_3 relative to 1900 are derived from the IPCC TAR (0xComp) models⁴. These values represent the radiative forcing due to changes in tropospheric ozone over the twenty-first century. They are added to the IPCC mean radiative forcing of 0.38 W m^{-2} (representing the period between pre-industrial to 2000)²⁸ from which a global mean radiative forcing value of 0.05 W m^{-2} (between 1850 and 1900) (ref. 27) is subtracted. The black triangle denotes the model mean, and the bars show the range across the models.

METHODS SUMMARY

Our approach is to modify net photosynthesis by a factor that accounts for plant ozone uptake and plant-specific sensitivities to ozone uptake. Ozone uptake is dependent on stomatal conductance, itself dependent on the photosynthetic rate in MOSES. The resulting equations are solved analytically to obtain a consistent solution for ozone uptake, stomatal conductance and ozone-modified net photosynthesis. Data from field observation^{16,20} are used to calibrate plant-ozone effects for the five plant functional types (PFTs) described by MOSES. A 'high' and 'low' parameterization is chosen for each PFT to represent uncertainty in the responses of different plant species to ozone deposition.

The Met Office's lagrangian tropospheric chemistry model STOCHEM¹⁹ was used to generate monthly mean surface ozone concentrations for the present day (2000) using emissions from the IIASA CLE scenario²⁸, and for the future (2100) using emissions from the SRES A2 scenario.

As described in ref. 7, this study uses the 0.5° resolution observational data set from the Climate Research Unit, which contains monthly temperature (mean and diurnal range), humidity, cloud cover and precipitation (amount and daily frequency). Empirical formulations are used to derive shortwave and longwave radiation from the Climate Research Unit data set. All monthly forcing data are regridded onto the HadCM3 $2.5^\circ \times 3.75^\circ$ grid and disaggregated to hourly data.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 September 2006; accepted 3 July 2007.

Published online 25 July 2007.

1. Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* 408, 184–187 (2000).
2. Friedlingstein, P. et al. Climate-carbon cycle feedback analysis: results from the C4MIP model intercomparison. *J. Clim.* 19, 3337–3353 (2006).
3. Gauss, M. et al. Radiative forcing since preindustrial times due to ozone change in the troposphere and the lower stratosphere. *Atmos. Chem. Phys.* 6, 575–599 (2006).

4. Gauss, M. *et al.* Radiative forcing in the 21st century due to ozone changes in the troposphere and the lower stratosphere. *J. Geophys. Res.* **108**, 4292, doi:10.1029/2002JD002624 (2003).
5. Ashmore, M. R. Assessing the future global impacts of ozone on vegetation. *Plant Cell Environ.* **28**, 949–964 (2005).
6. Karnosky, D. F. *et al.* Tropospheric O₃ modulates responses of temperate hardwood forests to elevated CO₂: a synthesis of molecular to ecosystem results from the Aspen FACE project. *Funct. Ecol.* **17**, 289–304 (2003).
7. Gedney, N. *et al.* Detection of a direct carbon dioxide effect in continental river runoff records. *Nature* **439**, 835–838 (2006).
8. Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 129–234 (Cambridge Univ. Press, Cambridge, UK, 2007).
9. Wang, X. & Mauzerall, D. L. Characterizing distributions of surface ozone and its impact on grain production in China, Japan and South Korea: 1900 and 2020. *Atmos. Environ.* **38**, 4383–4402 (2004).
10. Prather, M. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 239–287 (Cambridge Univ. Press, Cambridge, UK, 2001).
11. Felzer, B. S. *et al.* Future effects of ozone on carbon sequestration and climate change policy using a global biogeochemical model. *Clim. Change* **73**, 345–373, doi:10.1007/s10584-005-6776-4 (2005).
12. Field, C., Jackson, R. & Mooney, H. Stomatal responses to increased CO₂: implications from the plant to the global-scale. *Plant Cell Environ.* **18**, 1214–1255 (1995).
13. Karnosky, D. F. *et al.* Scaling ozone responses of forest trees to the ecosystem level in a changing climate. *Plant Cell Environ.* **28**, 965–981 (2005).
14. Volk, M. *et al.* Grassland yield declined by a quarter in 5 years of free-air ozone fumigation. *Glob. Change Biol.* **12**, 74–83, doi:10.1111/j.1365-2486.2005.01083.x (2006).
15. Percy, K. E. *et al.* New exposure-based metric approach for evaluating O₃ risk to North American aspen forests. *Environ. Pollut.* **147**, 554–566 (2007).
16. Pleijel, H. *et al.* Relationships between ozone exposure and yield loss in European wheat and potato — a comparison of concentration- and flux-based exposure indices. *Atmos. Environ.* **38**, 2259–2269 (2004).
17. Essery, R. L. H., Best, M. J., Betts, R. A., Cox, P. M. & Taylor, C. M. Explicit representation of sub-grid heterogeneity in a GCM land-surface scheme. *J. Hydrometeorol.* **4**, 530–543 (2001).
18. Cox, P. M. *et al.* The impact of new GCM land-surface physics on the GCM simulation of climate and climate sensitivity. *Clim. Dyn.* **15**, 183–203 (1999).
19. Sanderson, M. G., Jones, C. D., Collins, W. J., Johnson, C. E. & Derwent, R. G. Effect of climate change on isoprene emissions and surface ozone levels. *Geophys. Res. Lett.* **30**, 1936, doi:10.1029/2003GL017642 (2003).
20. Karlsson, P. E. *et al.* New critical levels for ozone effects on young trees based on AOT40 and simulated cumulative leaf uptake of ozone. *Atmos. Environ.* **38**, 2283–2294 (2004).
21. New, M., Hulme, M. & Jones, P. Representing twentieth-century space-time climate variability. Part II. Development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.* **13**, 2217–2238 (2000).
22. Denman, K. L. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 499–587 (Cambridge Univ. Press, Cambridge, UK, 2007).
23. Imhoff, M. L. *et al.* Global patterns in human consumption of net primary production. *Nature* **429**, 870–873 (2004).
24. Schröter, D. *et al.* Ecosystem service supply and vulnerability to global change in Europe. *Science* **310**, 1333–1337, doi:10.1126/science.1115233 (2005).
25. Nussbaum, S. & Fuhrer, J. Difference in ozone uptake in grassland species between open-top chambers and ambient air. *Environ. Pollut.* **109**, 463–471 (2000).
26. Sabine, C. L. *et al.* The oceanic sink for anthropogenic CO₂. *Science* **305**, 367–371 (2004).
27. Berntsen, T. K., Myhre, G., Stordal, F. & Isaksen, I. S. A. Time evolution of tropospheric ozone and its radiative forcing. *J. Geophys. Res.* **105**, 8915–8930 (2000).
28. Ramaswamy, V. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 350–416 (Cambridge Univ. Press, Cambridge, UK, 2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Gedney for technical support, and M. Sanderson for information on the STOCHEM fields used in this study; we acknowledge discussions with the aforementioned and with M. Ashmore, R. Betts, D. Hemming, O. Boucher and L. Mercado. We also thank A. Everitt for computer support. S.S. was supported by the UK Department for Environment, Food and Rural Affairs (DEFRA) Climate Prediction Programme. W.J.C. was supported by the MoD, and by DEFRA Air and Environment Quality Division, and C.H. by the UK Natural Environment Research Council.

Author Contributions P.M.C. developed the modification to MOSES to include ozone effects on photosynthesis and stomatal conductance; W.J.C. provided the projections of future changes in tropospheric ozone; C.H. developed the IMOGEN software that enabled the global simulations to be carried out; and S.S. calibrated the ozone effects model against data from manipulative field experiments, and carried out and analysed the global simulations. All four authors were involved in the drafting of the paper, although SS took the lead role.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.S. (stephen.sitch@metoffice.gov.uk).

METHODS

Modelling ozone effect in MOSES-TRIFFID. Our approach is to modify net photosynthesis, A_p , by an 'uptake of O_3 factor', F :

$$A = A_p F \quad (1)$$

From ref. 16,

$$F = 1 - aUO_{>FO3crit} \quad (2)$$

where F represents the fractional reduction of plant production, $UO_{>FO3crit}$ is the instantaneous leaf uptake of O_3 over a plant type specific threshold, $FO3crit$, in $nmol\ m^{-2}\ s^{-1}$. The fractional reduction of photosynthesis with O_3 uptake by leaves is represented by the plant type specific parameter, a (see Supplementary Table 1).

$$UO_{>FO3crit} = \max[(F_{O3} - FO3crit), 0.0] \quad (3)$$

From equations (2) and (3),

$$F = 1 - a \max[F_{O3} - FO3crit, 0.0] \quad (4)$$

F is dependent on the O_3 uptake rate by stomata over a critical (vegetation-dependent) threshold for damage. As an analogy of Ohm's law, the flux of O_3 to stomata, F_{O3} ($nmol\ O_3\ m^{-2}\ s^{-1}$), is given by,

$$F_{O3} = \frac{[O_3]}{R_a + \left[\frac{\kappa_{O3}}{g}\right]} \quad (5)$$

where $[O_3]$ is the molar concentration of O_3 at reference level ($nmol\ m^{-3}$), R_a is the aerodynamic and boundary layer resistance between leaf surface and reference level ($s\ m^{-1}$), g is the leaf conductance for H_2O ($m\ s^{-1}$), and $\kappa_{O3} = 1.67$ is the ratio of leaf resistance for O_3 to leaf resistance for water vapour. The uptake flux is dependent on the stomatal conductance, which is dependent on the photosynthetic rate in MOSES. Given that g is a linear function of photosynthetic rate, A , (equation (13) in ref. 18), from equation (1) it follows,

$$g = g_p F \quad (6)$$

where g_p is the leaf conductance in the absence of O_3 effects. The set of equations (4), (5) and (6) produces a quadratic in F that can be solved analytically.

To calibrate the ozone model, MOSES is first run from 1901 to 2002 assuming zero tropospheric ozone concentrations, and an initial pre-industrial equilibrium state. Year 2002 is repeated with geographically explicit tropospheric $[O_3]$ for 2002 from STOCHEM, using a first estimate of values for parameter a . Simulated half-hourly estimates of net primary productivity, NPP, and uptake of ozone, UO , are summed over the year. Relative annual yield, NPP/NPP_p (where NPP_p is the simulated NPP in the absence of plant ozone effects), is plotted against the cumulative annual uptake of ozone. The linear regression is compared with those derived from field observation^{16,20}. Values of parameter a are adjusted, and the procedure repeated until the linear regression through the simulation points replicates that obtained from field studies^{16,20} (Supplementary Fig. 2).

Data from field observation^{16,20} are used to calibrate plant-ozone effects for the five plant functional types (PFTs) described by MOSES: broadleaved trees, conifers, C_3 grasses, C_4 grasses and shrubs. A 'high' and 'low' parameterization is chosen for each PFT to represent species sensitive and less sensitive, respectively, to ozone effects. In the absence of data, the 'low' conifer parameterization is assumed to be 3.8 times less sensitive than the high parameterization (corresponding to the same ratio for the broadleaved trees).

These sensitivity parameters are typically based on the response of young trees to ozone exposure²⁰ mainly from open-top chamber experiments (OTC). In some climatic conditions (high radiation and temperature), the microclimate conditions in OTC may differ from ambient air²⁵. In the absence of observational data on plant-ozone uptake and ozone-induced yield reductions of trees in other life stages, we assume the response of young trees to be representative of all ages, and assume a constant ozone flux response of all plants throughout the growing period. The latter assumption may lead to an overestimate in the response of grasses to elevated ozone, as crops are known to exhibit variable flux response through the growing period. Threshold values, $FO3crit$, are taken at 1.6 and $5\ nmol\ m^{-2}\ s^{-1}$ for the woody and grass PFTs, respectively. Although a threshold of $5\ nmol\ m^{-2}\ s^{-1}$ implies a smaller O_3 dose for grasses, the gradient of the dose-response function, a , is larger, and therefore grasses may become more sensitive to ozone exposure than trees at high ozone concentrations. For shrubs we assume the same plant-ozone sensitivity as broadleaf trees. We prescribe some agricultural lands, fixed at present-day coverage throughout the simulations, in which grasslands are assumed to be dominant.

Future tropospheric ozone concentrations. The Met Office's lagrangian tropospheric chemistry model STOCHEM¹⁹ was used to generate monthly mean surface ozone concentrations for the present day (2000) using emissions from the IIASA CLE scenario²⁹, and for the future (2100) using emissions from the SRES A2 scenario.

Ozone levels for the pre-industrial period were also generated with this model⁴. Deposition fluxes of ozone and other trace gases were calculated using the stability of the boundary layer and fixed surface resistance values for land, sea and ice for each species. Biomass burning emissions were calculated using carbon emission data from ref. 30. The simulated surface ozone increases between 1900 and 2000 are mainly due to increased emissions of NO_x , but there have also been significant contributions from increases in hydrocarbon emissions associated with industrial activity in the northern mid-latitudes, and increasing biomass burning in the tropics.

The SRES A2 scenario for 2100 was chosen as this was the standard scenario used in the OxComp study presented in Chapter 4 of the IPCC Third Assessment Report¹⁰, although now some consider this to be a pessimistic scenario. Like all the SRES scenarios it assumes no emission control measures, and thus may be considered as an upper limit. The largest increases in future emissions predicted under A2 are from industrialization in India and China, along with population growth in the tropics.

For the future simulation the effect of climate change over land was to slightly decrease ozone concentrations by about 2 p.p.b. (ref. 29), except over polluted areas. This run did not have interactive isoprene, nor soil NO_x , which might be expected to increase surface ozone in a 2100 climate.

For each simulation (pre-industrial, present day and future), STOCHEM was integrated for 5 years and 4 months. For each calendar month, the average of the last five years of the run was calculated. Thus for each simulation 12 monthly fields of diurnal mean O_3 concentrations were generated. Values for intermediate years between those simulated were generated by linear interpolation for the appropriate month.

29. Dentener, F. D. *et al.* The global atmospheric environment for the next generation. *Environ. Sci. Technol.* **40**, 3586–3594 (2005).

30. van der Werf, G. R., Randerson, J. T., Collatz, G. J. & Giglio, L. Carbon emissions from fires in tropical and subtropical ecosystems. *Glob. Change Biol.* **9**, 547–562 (2003).

Talc-bearing serpentinite and the creeping section of the San Andreas fault

Diane E. Moore¹ & Michael J. Rymer¹

The section of the San Andreas fault located between Cholame Valley and San Juan Bautista in central California creeps at a rate as high as 28 mm yr^{-1} (ref. 1), and it is also the segment that yields the best evidence for being a weak fault embedded in a strong crust^{2–5}. Serpentinized ultramafic rocks have been associated with creeping faults in central and northern California^{6–8}, and serpentinite is commonly invoked as the cause of the creep and the low strength of this section of the San Andreas fault. However, the frictional strengths of serpentine minerals are too high to satisfy the limitations on fault strength, and these minerals also have the potential for unstable slip under some conditions^{9,10}. Here we report the discovery of talc in cuttings of serpentinite collected from the probable active trace of the San Andreas fault that was intersected during drilling of the San Andreas Fault Observatory at Depth (SAFOD) main hole in 2005. We infer that the talc is forming as a result of the reaction of serpentine minerals with silica-saturated hydrothermal fluids that migrate up the fault zone, and the talc commonly occurs in sheared serpentinite. This discovery is significant, as the frictional strength of talc at elevated temperatures is sufficiently low to meet the constraints on the shear strength of the fault, and its inherently stable sliding behaviour is consistent with fault creep¹¹. Talc may therefore provide the connection between serpentinite and creep in the San Andreas fault, if shear at depth can become localized along a talc-rich principal-slip surface within serpentinite entrained in the fault zone.

The SAFOD drillsite is located 14 km northwest of Parkfield in central California (Fig. 1), along the creeping section of the San

Andreas fault (SAF). In 2005, drilling successfully crossed the active trace of the SAF at $\sim 3 \text{ km}$ vertical depth¹², where the measured temperature is $\sim 112^\circ \text{C}$ (ref. 13). The drillhole terminated in sedimentary rocks of the Great Valley Sequence (K. McDougall, personal communication) east of the fault. Since then, a portion of the well casing has been actively deforming in response to creep on a fault strand¹². Serpentine was identified in X-ray diffraction patterns of cuttings¹⁴ collected at the eastern margin of the zone of active deformation (Supplementary Fig. 1). Aeromagnetic surveys¹⁵ indicate the presence of a flat-lying slab of serpentinite at $>3 \text{ km}$ depth on the northeast side of the fault (Fig. 1). This body may be $\geq 2 \text{ km}$ thick, and it abuts the fault for 50–60 km (ref. 15). The serpentinite slab is probably part of the Coast Range ophiolite, the oceanic basement on which the sediments of the Great Valley Sequence were deposited. Serpentinized ultramafic rock has a relatively low density compared to the overlying rock column, and a fault intersecting such a rock unit provides the pathway for the migration of serpentinite to shallower depths¹⁶. The Table Mountain serpentinite¹⁷ east of Parkfield (Fig. 1) is an extrusive body that formed as a result of the diapiric rise of low-density serpentinite from the deeply buried slab along faults that served as ‘fissure feeders’¹⁷. The serpentinite associated with the active trace in the SAFOD drillhole¹⁴ and outcrops of serpentinite^{18,19} fault gouge (Fig. 1) suggest that the same process is occurring along the SAF.

Serpentinite has been suggested as a possible cause of creep, because of its close association with creeping faults in central and northern California^{6–8}. The SAF creeping section coincides with the

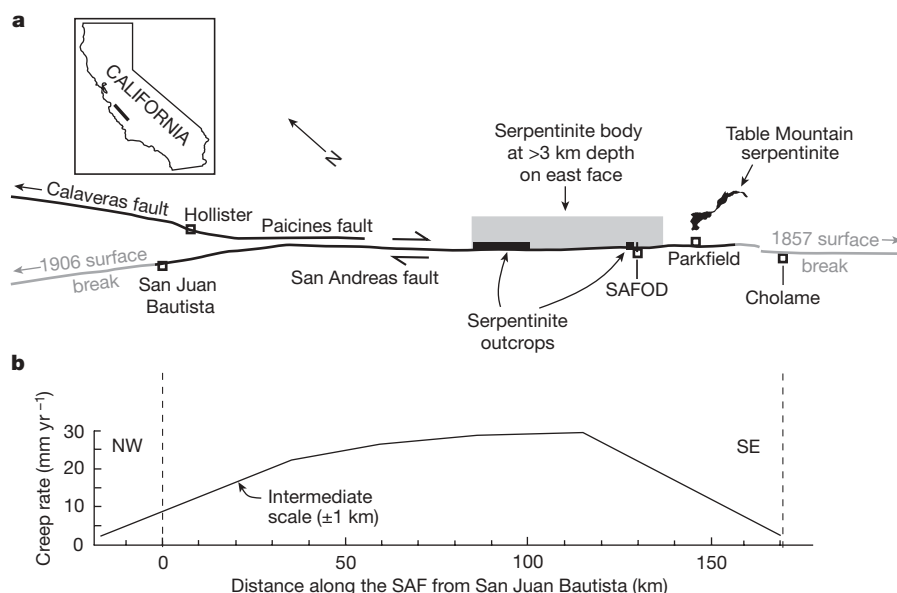


Figure 1 | Distribution of serpentinite along the SAF creeping section. **a**, The creeping section lies between areas of the fault that ruptured during great earthquakes in 1857 and 1906. Serpentine occurs in rare surface exposures of the fault^{18,19} and in the probable active trace of the fault encountered at $\sim 3 \text{ km}$ vertical depth in the SAFOD drillhole^{12,14}. The extrusive serpentinite at Table Mountain¹⁷ is derived from the same serpentinite body that abuts the fault on the northeast side at $>3 \text{ km}$ depth¹⁵. **b**, Recently updated creep rates measured at distances of 10 m to 1 km from the fault¹. Total offset rates along the San Andreas system in the creeping section are considered to be between 28 and 34 mm yr^{-1} (ref. 1).

¹US Geological Survey, 345 Middlefield Road, Mail Stop 977, Menlo Park, California 94025, USA.

The creeping section provides the best evidence for a weak SAF²⁻⁵. Because the creeping section is characterized by aseismic slip and microearthquakes, the apparent weakness of this segment cannot be explained through some dynamic weakening process accompanying a major earthquake, as it can for the locked sections. The restrictions on shear strength in the creeping SAF imposed by heat-flow^{22,23} and stress-orientation^{24,25} data are delimited in Fig. 2; also included are the frictional strengths of synthetic fault-gouge materials^{9-11,26} prepared by grinding and sieving rock or mineral separates. The strength experiments were conducted in a triaxial machine fitted with an internal furnace, at various combinations of temperature, confining and fluid pressure, and sliding velocity. For a given mineral to control the behaviour of the creeping section, it must be very weak as well as characterized by stable shear. The frictional properties of the serpentine minerals do not satisfy the weakness criterion and under certain conditions do not satisfy the stability criterion. The serpentine minerals lizardite and antigorite⁹ are not substantially weaker than granite²⁶ under hydrothermal conditions (Fig. 2). Chrysotile satisfies the heat-flow constraint to depths of ~3 km, but its strength increases substantially at greater depths^{9,10}. Furthermore, all three serpentine minerals show both velocity-weakening (strength

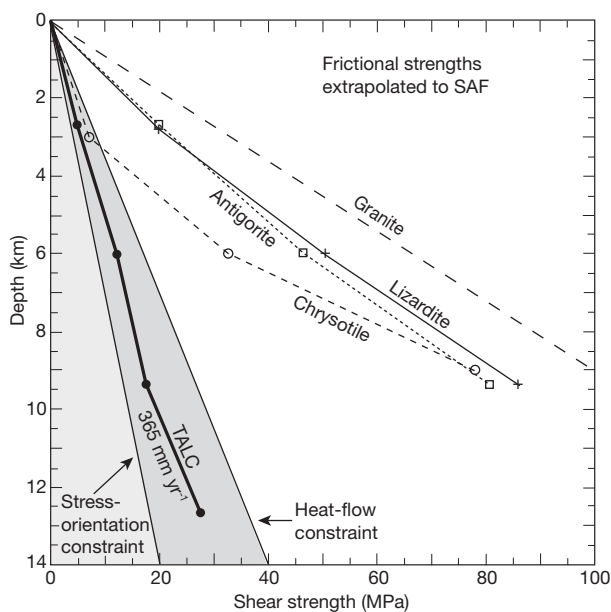
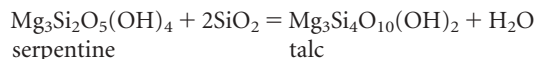


Figure 2 | Shear strength versus fault depth. Shaded fields indicate the constraints on the strength of the SAF based on heat-flow^{22,23} and stress-orientation^{24,25} investigations. Shear-strength data plotted for granite²⁶, serpentine minerals^{9–11} and talc¹¹ assume a temperature gradient of 30 °C km⁻¹ and a hydrostatic fluid-pressure gradient. At depths ≤ 3 km, both chrysotile and talc satisfy the heat-flow constraint, but chrysotile becomes substantially stronger at greater depths. The talc data represent a sliding velocity of 365 mm yr⁻¹. Given the characteristic velocity-strengthening behaviour of talc¹¹, its shear strength at ≤ 30 mm yr⁻¹ (Fig. 1) may be even lower.

The serpentinite contains numerous calcite- and some quartz-filled veins, possibly resulting from focused fluid flow within the fault zone. Talc replaces serpentine minerals along the vein walls (Fig. 3a, b), and it fills narrow cracks that extend into the serpentinite from the wider veins. Talc also forms along the foliation in sheared serpentinite grains (Fig. 3c, d). The talc-forming reaction is:



The SiO₂ comes from the dissolved silica content of heated ground water (Fig. 3a, c, d) and from quartz deposited metastably in veins (Fig. 3b). Talc is stable relative to the assemblage quartz + serpentine throughout the stability range of serpentine²⁷. The veins and shears with which talc is associated overprint all other textural features in the serpentinite grains, suggesting that the talc is of recent origin. Talc

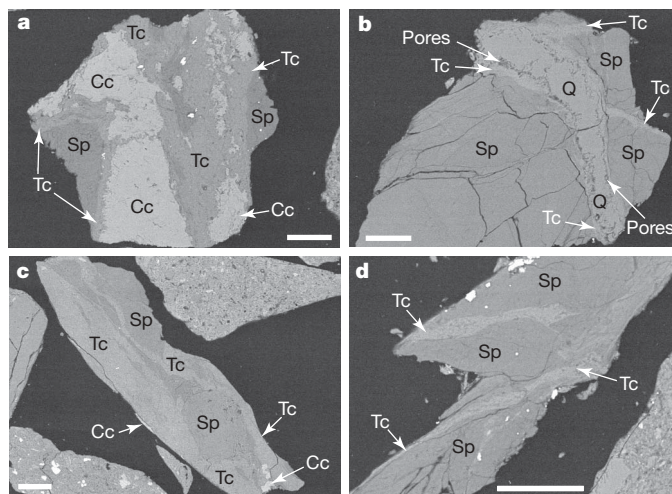


Figure 3 | Talc occurrences in serpentinite grains. Backscattered-electron SEM images of talc-bearing serpentinite grains from cuttings collected at 3,325 m MD. **a**, Talc (Tc) replacing serpentine minerals (Sp) adjacent to vein calcite (Cc). **b**, The reaction of serpentine and vein quartz (Q) to produce talc. The reaction results in a decrease in the volume of solid phases, consistent with the concentration of pores between the talc and quartz. **c**, Talc forming along the foliation in a sheared serpentinite grain. **d**, Talc in sheared serpentinite. Talc commonly appears at the edges of grains, perhaps because the serpentinite preferentially breaks along the weaker talc during drilling. Scale bars, 50 μ m.

is most abundant in the 3,325 m MD sample, which also contains nearly all of the sheared talc-bearing grains identified thus far. Talc is only a minor component of the serpentinite ($\leq 2\text{--}3\%$), although this soft mineral may have been preferentially lost during drilling (Fig. 3d).

Talc compositions (Supplementary Table 1) are consistent with those of talc in other low-temperature hydrothermal environments²⁷. The talc typically contains 4.0–5.5 wt% FeO and as much as 1.5 wt% NiO. In contrast, talc takes up only minor amounts of Al, and the Mg-rich smectite clay mineral saponite, with $\sim 5.0\text{--}7.5$ wt% Al_2O_3 (Supplementary Table 1), replaces serpentinite that has been mixed with feldspathic sediments.

The presence of talc in the active trace of the SAF is significant because talc has a very low shear strength in the temperature range 100–400 °C (Fig. 2). Talc may be the only mineral that can satisfy the conditions for a weak SAF over the entire depth range of the seismogenic zone without the need to invoke additional weakening mechanisms such as fluid overpressures. It is also characterized by inherently stable, velocity-strengthening behaviour¹¹. In rocks of appropriate composition, talc is stable at temperatures ranging from surficial to nearly 800 °C (ref. 27). The frictional strength of water-saturated smectite clay is comparable to that of talc at room temperature²⁸. However, the smectite clay saponite begins to break down at temperatures slightly above 100 °C (ref. 29), transforming to chlorite the water-saturated frictional strength of which²⁸ is close to that of lizardite. As with chrysotile (Fig. 2), saponite cannot explain the low apparent strength of the creeping section at depths greater than 3–4 km.

This discovery reinstates serpentinite as a possible explanation for fault creep, although indirectly through its association with talc; testing this hypothesis may prove challenging. The collection of core from the active trace of the SAF planned for 2007 at SAFOD may provide some answers, although, as noted previously, chrysotile and smectite clays have comparable frictional properties to talc at ~ 3 km depth. The small amount of talc found in the cuttings raises the question of whether enough talc could be present at greater depths to influence fault behaviour. However, along the Punchbowl fault, an exhumed former strand of the SAF in southern California, offset became extremely localized to a single fracture surface within a narrow (0.15–0.55 m) fault core³⁰. For such a fault geometry, only enough talc to line a fracture surface in serpentinite would be needed. Shear of laboratory samples of serpentinite and talc is typically highly localized^{9–11} along shear planes similar to the one in the Punchbowl fault. The talc-forming reaction should also be enhanced at depths > 3 km, because of faster reaction rates and the ability of higher-temperature ground waters to introduce larger amounts of dissolved silica to the serpentinite.

Received 6 March; accepted 3 July 2007.

- Titus, S. J., DeMets, C. & Tikoff, B. Thirty-five-year creep rates for the creeping segment of the San Andreas fault and the effects of the 2004 Parkfield earthquake: Constraints on alignment arrays, continuous global positioning system, and creepmeters. *Bull. Seismol. Soc. Am.* **96** (4B), S250–S268 (2006).
- Provost, A.-S. & Houston, H. Orientation of the stress field surrounding the creeping section of the San Andreas fault: Evidence for a narrow mechanically weak fault zone. *J. Geophys. Res.* **106**, 11373–11386 (2001).
- Hickman, S. & Zoback, M. Stress orientations and magnitudes in the SAFOD pilot hole. *Geophys. Res. Lett.* **31**, L15512, doi:10.1029/2004GL020043 (2004).
- Chéry, J., Zoback, M. D. & Hickman, S. A mechanical model of the San Andreas fault and SAFOD Pilot Hole stress measurements. *Geophys. Res. Lett.* **31**, L15513, doi:10.1029/2004GL019521 (2004).
- Williams, C. F., Grubb, F. V. & Galanis, S. P. Jr. Heat flow in the SAFOD pilot hole and implications for the strength of the San Andreas Fault. *Geophys. Res. Lett.* **31**, L15514, doi:10.1029/2003GL019352 (2004).
- Allen, C. R. in *Proc. Conf. on Geologic Problems of San Andreas Fault System* (eds Dickinson, W. R. & Grantz, A.) 70–80 (Stanford University Publications in the Geological Sciences Vol. 11, Stanford University, Stanford, California, 1968).
- Hanna, W. F., Brown, R. D. Jr, Ross, D. C. & Griscom, A. Aeromagnetic reconnaissance and generalized geologic map of the San Andreas fault between San Francisco and San Bernardino, California. *US Geol. Surv. Geophys. Investig. Map GP-815* (1972).
- Irwin, W. P. & Barnes, I. Effect of geologic structure and metamorphic fluids on seismic behavior of the San Andreas fault system in central and northern California. *Geology* **3**, 713–716 (1975).
- Moore, D. E., Lockner, D. A., Ma, S., Summers, R. & Byerlee, J. D. Strengths of serpentinite gouges at elevated temperatures. *J. Geophys. Res.* **102**, 14787–14801 (1997).
- Moore, D. E., Lockner, D. A., Tanaka, H. & Iwata, K. in *Serpentine and Serpentinities: Mineralogy, Petrology, Geochemistry, Ecology, Geophysics, and Tectonics* (ed. Ernst, W.G.) 525–538 (International Book Series Vol. 8, Geological Society of America, Boulder, Colorado, 2004).
- Moore, D. E. & Lockner, D. A. Comparative deformation behavior of minerals in serpentinitized ultramafic rock: Application to the slab-mantle interface in subduction zones. *Int. Geol. Rev.* **49**, 401–415 (2007).
- Zoback, M. D., Hickman, S. & Ellsworth, W. Overview of SAFOD Phases 1 and 2: Drilling, sampling and measurements of the San Andreas Fault zone at seismogenic depth. *Eos* **86** (Fall Meet. Suppl.), abstr. T23E-01 (2005).
- Williams, C. F., Grubb, F. V. & Galanis, S. P. Heat-flow measurements across the San Andreas fault near Parkfield, California — Preliminary results from SAFOD. *Eos* **87** (Fall Meet. Suppl.), abstr. S33B-0241 (2006).
- Solum, J. G. et al. Mineralogical characterization of protolith and fault rocks from the SAFOD main hole. *Geophys. Res. Lett.* **33**, L21314, doi:10.1029/2006GL027285 (2006).
- McPhee, D. K., Jachens, R. C. & Wentworth, C. M. Crustal structure across the San Andreas Fault at the SAFOD site from potential field and geologic studies. *Geophys. Res. Lett.* **31**, L12503, doi:10.1029/2003GL019363 (2004).
- Coleman, R. G. Petrologic and geophysical nature of serpentinites. *Geol. Soc. Am. Bull.* **82**, 897–918 (1971).
- Dickinson, W. R. Table Mountain serpentinite extrusion in California Coast Ranges. *Geol. Soc. Am. Bull.* **77**, 451–472 (1966).
- Rymer, M. J. Geologic map along a 12 kilometer segment of the San Andreas fault zone, southern Diablo Range, California (scale 1:12,000). *US Geol. Surv. Open-File Rep.* 81–1173 (1981).
- Rymer, M. J. et al. Surface fault slip associated with the 2004 Parkfield, California, earthquake. *Bull. Seismol. Soc. Am.* **96** (4B), S11–S27 (2006).
- Robbins, S. L. Complete Bouguer gravity, aeromagnetic, and generalized geologic map of the Hollister 15-minute quadrangle, California (scale 1:62,500). *US Geol. Surv. Geophys. Investig. Map GP-945* (1982).
- Galehouse, J. S. & Lienkaemper, J. J. Inferences drawn from two decades of alignment array measurements of creep on faults in the San Francisco Bay region. *Bull. Seismol. Soc. Am.* **93**, 2415–2433 (2003).
- Brune, J. N., Henyey, T. L. & Roy, R. F. Heat flow, stress and rate of slip along the San Andreas fault, California. *J. Geophys. Res.* **74**, 3821–3827 (1969).
- Lachenbruch, A. H. & Sass, J. H. Heat flow and energetics of the San Andreas fault zone. *J. Geophys. Res.* **85**, 6185–6223 (1980).
- Mount, V. S. & Suppe, J. State of stress near the San Andreas fault: Implications for wrench tectonics. *Geology* **15**, 1143–1146 (1987).
- Zoback, M. D. et al. New evidence for the state of stress on the San Andreas fault system. *Science* **238**, 1105–1111 (1987).
- Blanpied, M. L., Lockner, D. A. & Byerlee, J. D. Frictional slip of granite at hydrothermal conditions. *J. Geophys. Res.* **100**, 13045–13064 (1995).
- Evans, B. W. & Guggenheim, S. in *Hydrous Phyllosilicates (Exclusive of Micas)* (ed. Bailey, S.W.) 225–294 (Reviews in Mineralogy Vol. 19, Mineralogical Society of America, Washington DC, 1988).
- Moore, D. E. & Lockner, D. A. Crystallographic controls on the frictional behavior of dry and water-saturated sheet structure minerals. *J. Geophys. Res.* **109**, B03401, doi:10.1029/2003JB002582 (2004).
- Inoue, A. & Utada, M. Smectite-to-chlorite transformation in thermally metamorphosed volcanoclastic rocks in the Kamikita area, northern Honshu, Japan. *Am. Mineral.* **76**, 628–640 (1991).
- Chester, F. M. & Chester, J. S. Ultracataclastic structure and friction processes of the Punchbowl fault, San Andreas system, California. *Tectonophysics* **295**, 199–221 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.E.M. (dmoore@usgs.gov).

LETTERS

Female mate-choice drives the evolution of male-biased dispersal in a social mammal

O. P. Höner¹, B. Wachter¹, M. L. East¹, W. J. Streich¹, K. Wilhelm¹, T. Burke² & H. Hofer¹

Dispersal has a significant impact on lifetime reproductive success¹, and is often more prevalent in one sex than the other². In group-living mammals, dispersal is normally male-biased and in theory this sexual bias could be a response by males to female mate preferences, competition for access to females or resources, or the result of males avoiding inbreeding^{2–7}. There is a lack of studies on social mammals that simultaneously assess these factors and measure the fitness consequences of male dispersal decisions. Here we show that male-biased dispersal in the spotted hyaena (*Crocuta crocuta*) most probably results from an adaptive response by males to simple female mate-choice rules that have evolved to avoid inbreeding. Microsatellite profiling revealed that females preferred sires that were born into or immigrated into the female's group after the female was born. Furthermore, young females preferred short-tenured sires and older females preferred longer-tenured sires. Males responded to these female mate preferences by initiating their reproductive careers in groups containing the highest number of young females. As a consequence, 11% of males started their reproductive career in their natal group and 89% of males dispersed. Males that started reproduction in groups containing the highest number of young females had a higher long-term reproductive success than males that did not. The female mate-choice rules ensured that females effectively avoided inbreeding without the need to discriminate directly against close kin or males born in their own group, or to favour immigrant males. The extent of male dispersal as a response to such female mate preferences depends on the demographic structure of breeding groups, rather than the genetic relatedness between females and males.

Why is dispersal in most group-living mammals heavily biased towards males and which social groups offer males the best reproductive prospects? Answers to these questions are important because dispersal influences crucial components of lifetime reproductive success and is a major source of variance in fitness¹. In mammals with polygynous mating systems, females are assumed to incur higher costs from breeding with close relatives than males^{8,9}. These sexual asymmetries in costs are thought to cause sex-biased dispersal^{8,10}. High costs of inbreeding for females may favour female mate-choice towards immigrant males and discrimination against male kin^{3,11,12} and, in theory, female mate-choice can cause male-biased dispersal³.

We are unaware of any study that has assessed the impact on fitness of the decision by males about where to initiate their reproductive career and simultaneously evaluated the ultimate causes proposed for male-biased dispersal in social mammals. Here we tested whether male-biased dispersal in spotted hyaenas is driven by female mate-choice or by one of the other three main factors proposed to explain male-biased dispersal: male–male competition for access to females^{2,4,5}, inbreeding avoidance by males^{2,6}, or competition for

resources^{2,7}. We used ten years of detailed demographic data from the entire hyaena population (eight social groups) in the Ngorongoro Crater (hereafter referred to as 'Crater') in Tanzania, a habitat where processes of natural selection are still intact. To assess fitness benefits in terms of reproductive success of males after they initiated their reproductive career in a group we used microsatellite profiling of 426 offspring.

The spotted hyaena is a large carnivore that lives in social groups or 'clans' in which females socially dominate males¹³. Most but not all natal males disperse (that is, males leave the clan in which they were born and immigrate into a new clan)¹⁴, whereas female dispersal is very rare¹⁵. Immigrant male social status increases with length of tenure (time spent living in one group) because males observe strict social queueing conventions¹⁴. Because of the unusual anatomy of the female genitalia¹⁶, female cooperation is a prerequisite for intromission^{13,17}, and as a result females exercise considerable mate-choice¹⁸ and mate promiscuously in clans with numerous reproductively active males^{18,19}. Females are likely to incur far higher costs of inbreeding than males because only females care for offspring, lactating for an exceptionally long period and producing milk with a high protein, fat and energy content²⁰. Therefore, females would be expected to avoid breeding with close kin and to be choosier than males when selecting a mate^{9,21}.

In species such as the spotted hyaena, where females mate with several males in one oestrus cycle and males do not care for their young^{18,19}, females may not be able to distinguish their own father from other potential mates. A simple female mate-choice rule—'avoid males that were members of your group when you were born and favour males that were born into or immigrated into your group after your birth'—would reduce the chance of costly inbreeding of females with their father or with older brothers. Female Crater hyaenas generally adhered to this rule, choosing sires that were born into or immigrated into their clan after their birth more often than expected from the mean proportion of candidate males that fulfilled this requirement (Wilcoxon signed-rank test, $N = 64$ females of known age, $P = 0.0001$). Most females (81.3%, $N = 64$) always applied the rule; only a single female did not do so more than once. As a result, most litters (89.6%, $N = 134$) were sired by males that were born into or immigrated into the female's group after her birth (Fig. 1).

Young females (less than five years of age) produced litters sired by males with significantly shorter tenures than older females (Mann–Whitney U test, $U = 1,416.5$, $N_1 = 82$ litters by young females, $N_2 = 52$ litters by older females, $P = 0.001$). These results are consistent with the previously reported greater tolerance by young females of short-tenured males (with less than three years tenure) than longer-tenured males¹⁴, and the greater probability of offspring of young females being sired by short-tenured rather than

¹Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Strasse 17, D-10315 Berlin, Germany. ²Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.

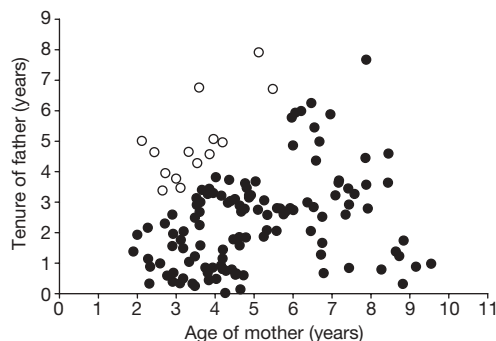


Figure 1 | The relationship between the age of the mother on the date of conception and the tenure of the father. ($N = 134$ litters.) Filled circles, litters sired by males that were born into or immigrated into the mother's clan after the mother's birth. Open circles, litters sired by males that were present when the mother was born.

long-tenured males (assortative mating)¹⁸. Given these female preferences, we would expect males to initiate their reproductive career in the clan with the highest number of young females, irrespective of whether this was their natal or another clan.

Spotted hyaena males are likely to assess potential dispersal destinations by undertaking short-term excursions into territories of other clans^{15,22}. Before natal adult males in the Crater initiated their reproductive career in their natal or a non-natal clan, they were frequently observed on excursions in territories of non-natal clans (mean proportion of 0.2 ± 0.02 of all sightings, $N = 114$ males, of which 62.3% were observed on such excursions). Furthermore, they were more often observed on such excursions than their twin sisters during the same period (Wilcoxon signed-rank test, $N = 20$ brother–sister twins, exact $P = 0.013$). Thus, males in the Crater are unlikely to be constrained in assessing potential dispersal destinations.

Of 142 males that were reared in Crater clans and reached adulthood, 114 males (80.3%) initiated their reproductive career in a clan on the Crater floor and 28 males (19.7%) died or dispersed elsewhere. Of the 114 males, 101 dispersed to a non-natal clan and 13 males (11.4%) initiated their reproductive career in their natal clan. Eleven males immigrated into Crater clans from elsewhere. For the 114 Crater-born males that initiated their reproductive career in a Crater clan, we assessed the key factors hypothesized to influence male dispersal (Table 1) for all eight clans on the Crater floor. As predicted, clan selection was influenced by the number of young females per group (Table 1), and males indeed preferred clans with the highest number of young females ($\chi^2 = 22.15$, degrees of freedom, d.f. = 1, $P < 0.001$; Fig. 2).

Males that initiated their reproductive tenure in clans with the highest number of young females obtained fitness benefits because male long-term reproductive success increased with the number of young females present at clan selection (stepwise backward regression, final model: $\ln(y) = -1.102 + 0.120x$, $F_{1,23} = 20.563$, $r^2 = 0.472$, $P < 0.001$; Fig. 3) after considering and removing from the

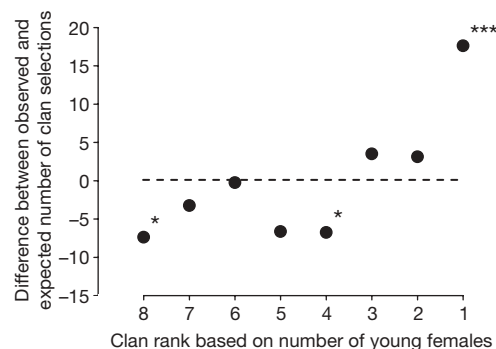


Figure 2 | Preference of male spotted hyaenas for clans with the highest number of young females. Clans were ranked in relation to the number of young female clan members on each date of clan selection by 114 males; in each case the clan with the highest number of young females had rank 1. * $P < 0.05$; *** $P < 0.001$.

model the annual rate of mortality of these females (full model: $t = -1.07$, $P = 0.298$). Furthermore, the long-term reproductive success of such males was higher than that of other males ($U = 8$, $N_1 = 9$ males that initiated their reproductive career in clans with the highest number of young females, $N_2 = 16$ males that initiated their reproductive tenure in clans that did not contain the highest number of young females, exact $P < 0.0001$; Fig. 4).

Males that initiated reproductive activity in the clan with the highest number of young females were likely to secure long-term access to numerous mating partners because survivorship of these females was above 75% during the first six years of male tenure. Thus, a judicious clan selection provides males with a high number of females with which they can develop long-term 'friendly' associations as both male and female tenure increases—a male tactic actively preferred by females that promotes male reproductive success¹⁸.

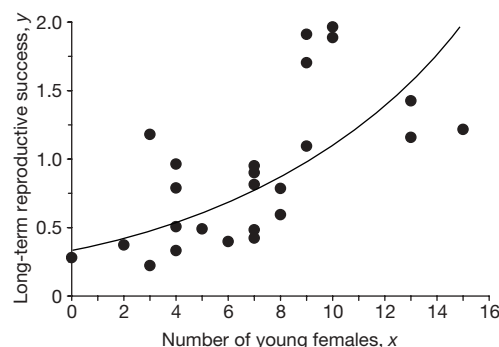


Figure 3 | The influence of the number of young female clan members at clan selection on the long-term reproductive success of male spotted hyaenas. Long-term reproductive success was the mean number of cubs produced per year of tenure for 25 males with a minimum tenure of four years in a clan. The line shows regression of long-term reproductive success by number of young females, $y = 0.332 \times e^{0.120x}$.

Table 1 | Test of predictions from the main hypotheses for the evolution of male-biased dispersal

Hypothesis	Variable predicted to influence the likelihood of males to select a clan	Model coefficient	Standard error	t-ratio	P
Avoidance of competition with other males for access to females Female mate-choice	Intensity of male–male competition*	0.001	0.029	0.024 (0.122*)	0.981 (0.903*)
	Number of females most likely to breed with males ('young females' as defined in Methods)	0.072	0.034	2.102 (2.064*)	0.036 (0.039*)
Avoidance of breeding with close female relatives	Number of unrelated adult females with relatedness of < 0.5	0.005	0.024	0.224 (0.298*)	0.823 (0.766*)
Avoidance of competition for resources	Number of main prey animals per adult or yearling spotted hyaena	0.001	0.001	0.464 (0.456*)	0.643 (0.649*)

Discrete choice regression model with the identity of the clan selected by 114 males as dependent variable; log-likelihood of whole model = -229.988 .

* 'Intensity of male–male competition' refers to the number of reproductively active natal males plus immigrant males. In an alternative model (t-ratios and P values given in parentheses), the number of reproductively active males per adult female was chosen instead.

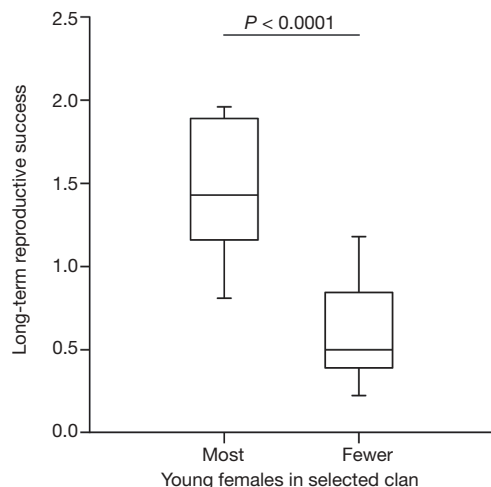


Figure 4 | The fitness benefits of male spotted hyaenas that selected clans with the highest number of young females. Long-term reproductive success calculated as in Fig. 3. The box indicates the interquartile range around the median (line inside the box), and the vertical error bars represent values plus or minus 1.5 times the interquartile range.

As male tenure in a clan increased, the number of females in the clan that were young at the time the males initiated their reproductive career in the clan declined because of female mortality; after seven years of tenure only 59.3% and after eight years only 45.8% of these females remained alive. This may explain the decrease in reproductive performance of long-tenured males towards the end of their tenures^{18,19}, and (secondary) dispersal of 16.7% of Crater males with tenures exceeding six years, despite the fact that these males had obtained a high social status in the male hierarchy.

There was little evidence that females produced offspring sired by close male relatives, thereby risking fitness costs of inbreeding. Only four of 426 cubs in two of 309 litters (0.6%) produced by one of 110 females (0.9%) resulted from daughter–father matings. Females that conceived when their father was a member of their clan produced only two of 88 litters with their father. None of the five litters that mothers conceived when their sons were reproductively active in their clan resulted from mother–son matings. None of the males that were reproductively active in their natal clan had a sister that conceived during their tenure, so breeding between females and their brothers could not occur.

Males did not appear to avoid the chance of breeding with close female relatives, because clan selection was independent of the number of unrelated females in a clan (Table 1). Furthermore, of 76 males that consorted with (shadowed¹⁴) females, the 13 males that had the opportunity to shadow daughters did not shadow daughters less often than expected from the mean proportion of daughters in the pool of adult females (Wilcoxon signed-rank test, exact $P = 0.542$).

Males did not select clans with respect to the likely level of male–male competition because clan selection was independent of two measures of competition: the total number of male competitors and the number of male competitors per adult female (Table 1). Males thus did not prefer clans with short male social queues, a result consistent with the idea that when the number of potential mating partners available to males in a long queue is greater than that in a shorter queue, males benefit more by joining longer queues²³.

There was also no evidence that clan selection was influenced by competition for food because selection was independent of the per capita number of main prey animals in a clan territory (Table 1).

Our findings suggest that female mate-choice is the main factor determining the clan in which males initiate their reproductive career. Males that responded best to the observed female mate preferences had the highest long-term reproductive success. We

conclude that female mate-choice represents a sufficient cause for the evolution of sex-biased dispersal in social mammals.

The observed female preferences were simple mate-choice rules that radically reduced the chance of costly inbreeding for females. These rules do not require direct kin discrimination, nor indirect location-based kin discrimination (such as preference for immigrant males or discrimination against natal males)^{3,11,12}, both important parameters in theoretical models that seek to explain male dispersal in social mammals³. Instead, they are consistent with indirect time-based kin discrimination cues²⁴. This means that an intrinsically demographic property—fluctuations in the number of young females in different clans—can lead to male dispersal in the majority of cases. Thus, changes in the demographic structure of groups will alter the likelihood of males dispersing, and the demographic structure of a group in relation to other groups will set the level of emigration from and immigration into that group.

METHODS SUMMARY

Study area and groups. All approximately 370 hyaenas of the eight Crater clans were individually known¹⁵ and studied between April 1996 and April 2006. Natal adult males that attempted to mate with or shadowed females from their natal clan or that excluded competing males from access to a female²⁵ were termed ‘reproductively active natal males’. Immigrant males were considered members of the new clan if they initiated non-aggressive interactions with its members over a period of at least three months. Date of clan selection was the date of first sighting in the new clan’s territory (immigrant males) or of first observation of mating, shadowing, or defending (reproductively active natal males). Male tenure was calculated as the number of days from the date of clan selection until the date of the event in question.

Social status, clan selection and paternity analysis. To test whether males preferred to initiate their reproductive career in the clan with the highest number of young females, the eight clans (including each male’s natal clan) were placed in a linear rank order on each date a male selected a clan. Rank 1 was the clan with the highest, and rank 8 the clan with the lowest number of young females. This expected pattern of clan selection was compared with the observed pattern by calculating Manly’s standardized selection ratio²⁶ $B = \text{Chesson's } \alpha$. Number of young females included females that were between one and five years of age¹⁸. Relatedness between individuals was calculated from known pedigrees based on genetic paternity analyses; first-degree relatives were referred to as ‘closely related’. Paternity analyses were based on amplification of six highly polymorphic microsatellite loci using genetic material from 575 Crater individuals collected as previously described¹⁸. Results are quoted as means \pm standard error, and probabilities are for two-tailed tests.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 March; accepted 21 June 2007.

1. Alberts, S. C. & Altmann, J. Balancing costs and opportunities: dispersal in male baboons. *Am. Nat.* **145**, 279–306 (1995).
2. Greenwood, P. J. Mating systems, philopatry and dispersal in birds and mammals. *Anim. Behav.* **28**, 1140–1162 (1980).
3. Lehmann, L. & Perrin, N. Inbreeding avoidance through kin recognition: choosy females boost male dispersal. *Am. Nat.* **162**, 638–652 (2003).
4. Dobson, F. S. Competition for mates and predominant juvenile male dispersal in mammals. *Anim. Behav.* **30**, 1183–1192 (1982).
5. Moore, J. & Ali, R. Are dispersal and inbreeding avoidance related? *Anim. Behav.* **32**, 94–112 (1984).
6. Cockburn, A., Scott, M. P. & Scotts, D. J. Inbreeding avoidance and male-biased natal dispersal in *Antechinus* spp. (Marsupialia: Dasyuridae). *Anim. Behav.* **33**, 908–915 (1985).
7. Johnson, M. L. & Gaines, M. S. Evolution of dispersal: theoretical models and empirical tests using birds and mammals. *Annu. Rev. Ecol. Syst.* **21**, 449–480 (1990).
8. Smith, R. H. On selection for inbreeding in polygynous animals. *Heredity* **43**, 205–211 (1979).
9. Parker, G. A. in *Mate Choice* (ed. Bateson, P. P. G.) 141–166 (Cambridge Univ. Press, Cambridge, 1983).
10. Parker, G. A. in *Sexual Selection and Reproductive Competition in Insects* (eds Blum, M. S. & Blum, N. A.) 123–166 (Academic Press, New York, 1979).
11. Pusey, A. E. & Packer, C. in *Primate Societies* (eds Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W. & Struhsaker, T. T.) 250–266 (Univ. Chicago Press, Chicago, 1986).
12. Keane, B. Dispersal and inbreeding avoidance in the white-footed mouse, *Peromyscus leucopus*. *Anim. Behav.* **40**, 143–152 (1990).

13. Kruuk, H. *The Spotted Hyena. A Study of Predation and Social Behavior* (Univ. Chicago Press, Chicago, 1972).
14. East, M. L. & Hofer, H. Male spotted hyenas (*Crocuta crocuta*) queue for status in social groups dominated by females. *Behav. Ecol.* **12**, 558–568 (2001).
15. Höner, O. P., Wachter, B., East, M. L., Runyoro, V. A. & Hofer, H. The effect of prey abundance and foraging tactics on the population dynamics of a social, territorial carnivore, the spotted hyena. *Oikos* **108**, 544–554 (2005).
16. Matthews, L. H. Reproduction in the spotted hyaena, *Crocuta crocuta*, (Erleben). *Phil. Trans. R. Soc. Lond. B* **230**, 1–78 (1939).
17. East, M. L., Hofer, H. & Wickler, W. The erect 'penis' is a flag of submission in a female-dominated society: greetings in Serengeti spotted hyaenas. *Behav. Ecol. Sociobiol.* **33**, 355–370 (1993).
18. East, M. L., Burke, T., Wilhelm, K., Greig, C. & Hofer, H. Sexual conflicts in spotted hyenas: male and female mating tactics and their reproductive outcome with respect to age, social status and tenure. *Proc. R. Soc. Lond. B* **270**, 1247–1254 (2003).
19. Engh, A. L. *et al.* Reproductive skew among males in a female-dominated mammalian society. *Behav. Ecol.* **13**, 193–200 (2002).
20. Hofer, H. & East, M. L. in *Serengeti II. Dynamics, Management, and Conservation of an Ecosystem* (eds Sinclair, A. R. E. & Arcese, P.) 332–363 (Elsevier, Chicago, 1995).
21. Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 136–179 (Aldine, Chicago, 1972).
22. Smale, L., Nunes, S. & Holekamp, K. E. in *Advances in the Study of Behavior* Vol. 26 (eds Slater, P. J. B., Rosenblatt, J. S., Milinski, M. & Snowdon, C. T.) 181–250 (Academic Press, London, 1997).
23. Kokko, H. & Sutherland, W. J. Optimal floating and queuing strategies: consequences for density dependence and habitat loss. *Am. Nat.* **152**, 345–366 (1998).
24. Sherman, P. W., Reeve, H. K. & Pfennig, D. W. in *Behavioural Ecology—An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 69–96 (Blackwell Science, Oxford, 1997).
25. East, M. L. & Hofer, H. Loud calling in a female-dominated mammalian society: II. Behavioural contexts and functions of whooping of spotted hyaenas, *Crocuta crocuta*. *Anim. Behav.* **42**, 651–669 (1991).
26. Manly, B. F. J., McDonald, L. L. & Thomas, D. L. *Resource Selection by Animals. Statistical Design and Analysis for Field Studies* (Chapman & Hall, London, 1993).

Acknowledgements We thank the Tanzania Commission for Science and Technology for permission to conduct the study, the Tanzania Wildlife Research Institute, the Ngorongoro Conservation Area Authority, A. Francis, L. Kimaay, T. Ndoto, G. Orio, H. Richner, D. Thierer, C. Trout, L. Trout, C. Voigt and W. Wickler for their assistance and suggestions. This study was financed by the Leibniz Institute for Zoo and Wildlife Research, the Fritz-Thyssen-Stiftung, the Stifterverband der deutschen Wissenschaft, the Max Planck Society, the German Academic Exchange Service (DAAD) and the Messerli Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.P.H. (hoener@izw-berlin.de).

METHODS

Study area and population. The floor of the Ngorongoro Crater in northern Tanzania covers 250 km² and is inhabited by eight spotted hyaena clans with between 24 and 65 members each that defended territories of 24 ± 4 km² (ref. 15). The Crater population is linked to the neighbouring Serengeti population by individual movements and gene flow²⁷. Both populations contain reproductively active natal males and both have a similar incidence of inbreeding (Crater: 0.9%; Serengeti¹⁸: 0.8%) and expected heterozygosity (Crater: 0.824, this work; Serengeti¹⁸: 0.856). Sex, age and social status of individuals were determined as previously described¹⁵. Individuals less than 12 months of age were classified as cubs, those aged between 12 and 24 months as yearlings, and those 24 months of age or older as adults. The date of conception was calculated from litter birth dates on the basis of a gestation period of 110 days¹⁶.

Selection of clans. We assessed hypotheses for the evolution of male-biased dispersal using a discrete choice (multinomial logistic) regression model²⁸ by asking which of four variables predicts the clan in which males started their reproductive career (this may have been their natal clan). The variables were (1) intensity of male–male competition, (2) number of young females, (3) number of unrelated females, and (4) mean number of main prey animals per hyaena (adults and yearlings) on the dates of clan selection (Table 1). Intensity of male–male competition was the length of the male social queue (that is, the number of reproductively active natal males plus immigrant males), or the number of reproductively active males per adult female. In spotted hyaenas, the length of the male queue may be the more appropriate measure of male–male competition because males need to build friendly relationships with females to reproduce and queue for social status, and as a result, levels of aggression between males are low¹⁴.

From the perspective of each male, young females in non-natal clans were those between one and less than five years of age on the date of clan selection, since recent immigrant males rarely have contact with female cubs less than 12 months of age²³. Young females in his natal clan were those that were born before the male's birth and less than five years of age. The number of unrelated females was all adult females with a coefficient of relatedness $r < 0.5$. The mean number of main prey animals per hyaena was determined from data on mean main prey density and territory size¹⁵ divided by the mean clan size (adults and yearlings).

Genetic analysis and survivorship of females. Methods for the collection and processing of genetic material for paternity analysis have been previously described^{18,29}. Microsatellite loci were typed for 575 Crater individuals including 434 offspring (65.2% of all offspring born during the study period). Paternity was assessed using maximum-likelihood methods as implemented in Cervus³⁰. All immigrant and reproductively active natal males that were clan members when a litter was conceived were considered to be putative fathers. The mean proportion of candidate males that were typed was 0.979; for 386 (88.9%) offspring all candidate males were typed. Hence, for 426 (98.2%) of the 434 offspring from which DNA was isolated, paternity was determined with 95% confidence. The mean expected heterozygosity was 0.824, total exclusionary power was 0.999, the mean proportion of individuals typed was 0.992, and the error rate was 0.0052 and was set at 1.0%. The survivorship of young females was calculated as the mean proportion of young females present at clan selection that survived to the end of each year of male tenure.

Statistical analysis. Nonparametric tests, the discrete choice regression model and the stepwise regression model were performed using Systat 11.0 (Systat Software Inc.). For the stepwise regression model, natural-logarithm transformation was applied to the dependent variable to satisfy the requirement of normal distribution of residuals as judged by the Lilliefors test. The significance of Wilcoxon signed-rank and Mann–Whitney U tests with sample sizes below 30 were based on exact P -values calculated with StatXact 7.0 (Cytel Inc.).

27. Albert, R. *Genstruktur und Genfluß in ausgewählten Populationen der Tüpfelhyäne (Crocuta crocuta)*. PhD thesis, Freie Universität Berlin (2002).

28. McFadden, D. in *Frontiers in Econometrics* (ed. Zarembka, P.) 105–142 (Academic Press, New York, 1974).

29. Wilhelm, K. et al. Characterization of spotted hyena, *Crocuta crocuta*, microsatellite loci. *Mol. Ecol. Notes* 3, 360–362 (2003).

30. Marshall, T. C., Slate, J., Kruuk, L. E. B. & Pemberton, J. M. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7, 639–655 (1998).

LETTERS

Correlation between neural spike trains increases with firing rate

Jaime de la Rocha^{1*}, Brent Doiron^{1,2*†}, Eric Shea-Brown^{1,2}, Krešimir Josić³ & Alex Reyes¹

Populations of neurons in the retina^{1–3}, olfactory system⁴, visual⁵ and somatosensory⁶ thalamus, and several cortical regions^{7–10} show temporal correlation between the discharge times of their action potentials (spike trains). Correlated firing has been linked to stimulus encoding⁹, attention¹¹, stimulus discrimination⁴, and motor behaviour¹². Nevertheless, the mechanisms underlying correlated spiking are poorly understood^{2,3,13–20}, and its coding implications are still debated^{13,16,21,22}. It is not clear, for instance, whether correlations between the discharges of two neurons are determined solely by the correlation between their afferent currents, or whether they also depend on the mean and variance of the input. We addressed this question by computing the spike train correlation coefficient of unconnected pairs of *in vitro* cortical neurons receiving correlated inputs. Notably, even when the input correlation remained fixed, the spike train output correlation increased with the firing rate, but was largely independent of spike train variability. With a combination of analytical techniques and numerical simulations using ‘integrate-and-fire’ neuron models we show that this relationship between output correlation and firing rate is robust to input heterogeneities. Finally, this overlooked relationship is replicated by a standard threshold-linear model, demonstrating the universality of the result. This connection between the rate and correlation of spiking activity links two fundamental features of the neural code.

How do cortical cells transform correlation between their synaptic currents into correlation between their output spike trains? We addressed this question by studying pairwise spike train correlations, a strategy that can capture the full statistical structure of a neural network^{2,3}. Correlated fluctuating currents resembling synaptic activity²³ were injected into the somata of 20 unconnected cortical neurons (see Methods). The input current to cell i ($i = 1, 2$) was:

$$I_i = \mu_i + \sigma_i \left(\sqrt{1-c} \xi_i(t) + \sqrt{c} \xi_c(t) \right) \quad (1)$$

where μ_i is the temporal average of the current. The next term represents gaussian fluctuations with a temporal structure consistent with that *in vivo*²³, and was composed of two weighted factors: ξ_i , which was independent for each cell, and ξ_c , which was common to all cells (Fig. 1a). The input correlation coefficient, c ($0 \leq c \leq 1$), set the relative weight of the shared fluctuations, whereas σ_i set the variance of the total input current. These parameters were adjusted so that the variability and covariability of the membrane potentials²⁴ and spike trains^{7,8} were similar to those observed *in vivo* (Fig. 1b). Typical spike train cross-correlation functions had positive central peaks with short timescales (~ 20 ms; Fig. 1c, left), indicative of spike-time synchrony⁵.

To quantify spike train correlation between two cells, we computed the correlation coefficient^{1,7,8,10} of the spike counts, n_1 and

n_2 , over a sliding window of length T :

$$\rho_T = \frac{\text{Cov}(n_1, n_2)}{\sqrt{\text{Var}(n_1)\text{Var}(n_2)}} \quad (2)$$

where Cov is the covariance and Var the variance. The correlation coefficient ρ_T is a dimensionless quantity ranging between 0 for independent and 1 for fully correlated spike trains. For large T , ρ_T saturates to a value $\rho^{8,10}$ (see Methods).

As expected, ρ_T increased with $c^{14,15,17}$ (Fig. 1d); however, ρ_T was always less than c , indicating that the correlation in the input currents bounds the correlation between the output spikes. Our central and more surprising finding was that ρ_T increased with the output firing rate v when c was fixed (Fig. 1e). We delivered currents with a fixed $c = 0.5$ and various values of μ and σ , thus eliciting spike trains with a range of firing rates and inter-spike interval coefficients of variation (CV). Spike trains with similar v were paired to compute ρ_T . Despite ρ_T increasing strongly with rate it did not depend on the CV (inset Fig. 1e). The same increasing trend holds for all T values examined, and occurred across several identified cell types (see Supplementary Information).

The correlation–rate relationship (Fig. 1e) might be expected owing to an increased probability of spikes occurring close in time at high rates. The coefficient ρ_T corrects for this spurious rate-dependence of correlation, as illustrated by the following example of two cells receiving inputs constructed from three independent Poisson trains x_1 , x_2 and x_c with rates $(1-c)v$, $(1-c)v$ and cv , respectively (Fig. 1f). For a simple model that converts every input spike into an output spike, the output trains $y_1 = x_1 + x_c$ and $y_2 = x_2 + x_c$ have rate v and a correlation that depends on the rate of the common train x_c (Fig. 1f, red spikes). The correlation measured as the average product $\langle n_1 n_2 \rangle = cvT + (vT)^2$ depends on v through: (1) the baseline chance correlation $(vT)^2$, which gives a non-zero correlation even in the case $c = 0$; (2) the scaling of the first term by v , which gives the absolute number of synchronous spikes. The coefficient ρ_T (equation (2)) corrects for these v -dependencies through baseline subtraction and appropriate normalization: $\rho_T = \frac{\langle n_1 n_2 \rangle - \langle n_1 \rangle \langle n_2 \rangle}{\sqrt{(\langle n_1^2 \rangle - \langle n_1 \rangle^2)(\langle n_2^2 \rangle - \langle n_2 \rangle^2)}}$. Specifically, noting $\langle n_i \rangle = vT$ and $\langle n_i^2 \rangle = vT + (vT)^2$ yields $\rho_T = cvT/vT = c$, thus showing how ρ_T does not have a built-in dependence on v . Therefore the correlation–rate relationship exhibited by the data (Fig. 1e) is a priori unexpected, and instead must follow from an input–output property of spiking neurons.

To analyse further the correlation–rate relationship we replaced the *in vitro* neurons with a pair of leaky ‘integrate-and-fire’ (LIF) model neurons²⁵ (see Methods). We fixed μ and σ , and performed simulations of the LIF neuron pair to compute $\rho(c)$. The coefficient c ranged from 0 to 0.3, thus yielding ρ values that are comparable to

¹Center for Neural Science, New York University, New York 10003, USA. ²Courant Institute of Mathematical Sciences, New York University, New York 10012, USA. ³Department of Mathematics, University of Houston, Houston, Texas 77204, USA. [†]Present address: Department of Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. *These authors contributed equally to this work.

those measured *in vivo*^{7,8,11}. In this range $\rho(c)$ was approximately linear (Fig. 2a, b), allowing us to use perturbation techniques²⁶ to obtain an expression for $\rho(c)$ (see Supplementary Information):

$$\rho \approx S(\mu, \sigma) c = \frac{\sigma^2 \left(\frac{dv}{d\mu} \right)^2}{CV^2 v} c \quad (3)$$

We define the slope $S(\mu, \sigma)$ relating ρ to c as correlation susceptibility. Here $dv/d\mu$ is the derivative of the spike-rate transfer function (that is, the relationship between firing rate, $v(\mu, \sigma)$, and the mean injected current). Equation (3) accurately predicts the transformation $\rho(c)$ obtained via simulations (Fig. 2a, b, dashed lines).

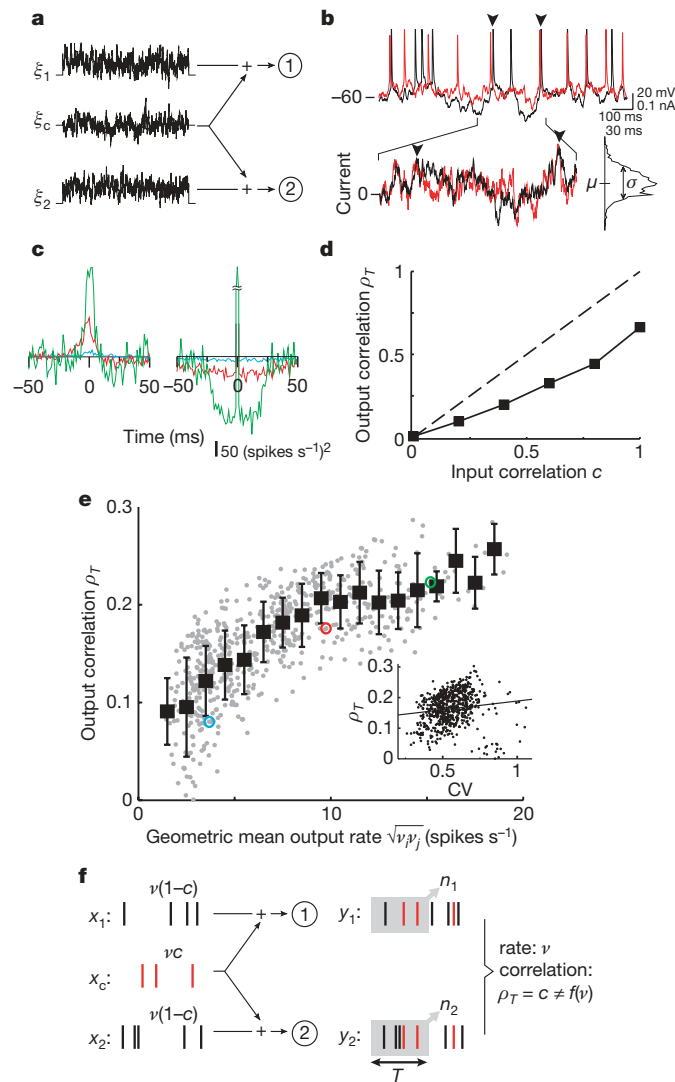


Figure 1 | Relationship between output spike correlation and rate in *in vitro* cells. **a**, Twenty neurons from cortical slices ($n = 11$) were stimulated with gaussian currents (filtered with a time constant of 5 ms), with both common and independent components. **b**, Fluctuating current injection (bottom) often evoked nearly synchronous spikes (top arrowheads). Red and black traces are membrane potentials from cells 1 and 2, respectively. **c**, Typical cross- (left) and auto-correlation (right) functions from a pair of cells firing at low (blue), medium (red) and high (green) rate. **d**, A typical ρ_T versus c for a pair of cells. **e**, The correlation ρ_T in a population ($n = 123$ pairwise comparisons) increases with $\sqrt{v_1 v_2}$ and does not co-vary with $\sqrt{CV_1 CV_2}$ (inset, linear regression, $r = 0.13$); c was fixed at 0.5. Squares represent mean and s.d. of the population, and coloured circles the examples shown in **c**. **f**, A simple cell model that generates an output spike for every input spike. The correlation ρ_T is fixed at c , and is independent of v . Throughout we used $T = 40$ ms.

As with the experiments, we varied μ or σ while keeping c fixed to obtain different values of v . A plot of ρ versus v when μ (or σ) was increased while keeping σ (or μ) fixed (inset of Fig. 2c) shows that ρ rapidly increased in the physiological range of 0–50 spikes s^{-1} and asymptoted to c at high rates (Fig. 2c). Notably, the correlation–rate curves obtained by increasing μ or σ were nearly identical, even though voltage traces, correlation functions and CVs were very different (Fig. 2c). To show that these results did not depend on specific μ and σ values, we used equation (3) to sweep a large region of the μ – σ parameter space. The resultant ρ – v values were confined to a narrow band around the stereotyped curve (Fig. 2c, shaded areas). This allowed us to rewrite equation (3) in an approximate form in which the susceptibility, S , depends only on the output rate and not on the input variables μ and σ :

$$\rho \approx S(v) c \quad (4)$$

The empirically measured ρ_T also increased with v for all T examined; however, the approximation that S depends only on v (equation (4)) becomes progressively less accurate as T decreases (see Supplementary Information).

To determine whether the correlation–rate relationship emerges in a less controlled and more biologically relevant setting, we performed simulations of a two-layer network where neurons from layer 1 were randomly connected to neurons in layer 2 (Fig. 3a). Layer 1 consisted

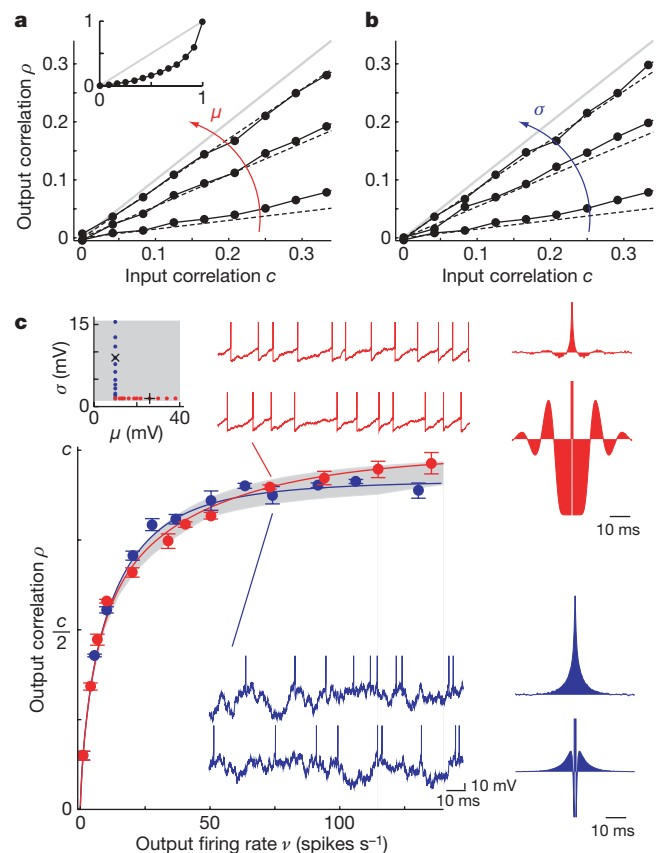


Figure 2 | The correlation–rate relationship in an integrate-and-fire neuron model. **a**, ρ versus c from simulations (solid line) and theory (dashed line, from equation (3)). Increases in either μ (**a**) or σ (**b**) increase the slope of $\rho(c)$. Grey lines are the diagonal $\rho = c$. Inset shows $\rho(c)$ for all c . **c**, As μ (red) or σ (blue) were varied following the paths drawn in the inset, ρ and v co-varied as shown. Note that this relationship holds only when $\rho(c)$ is linear (**a** and **b**). Simulations (dots) and theory (solid line) show a dependence of ρ and v along a stereotyped curve. The shaded area shown in μ – σ space (inset) was swept using equation (3) and yielded the shaded area around the ρ – v curves. Voltage traces, cross- and auto-correlation functions (upper and lower panels, respectively) are illustrated for two sets of (μ, σ) values (see + and \times in inset). We used $\tau_m = 10$ ms.

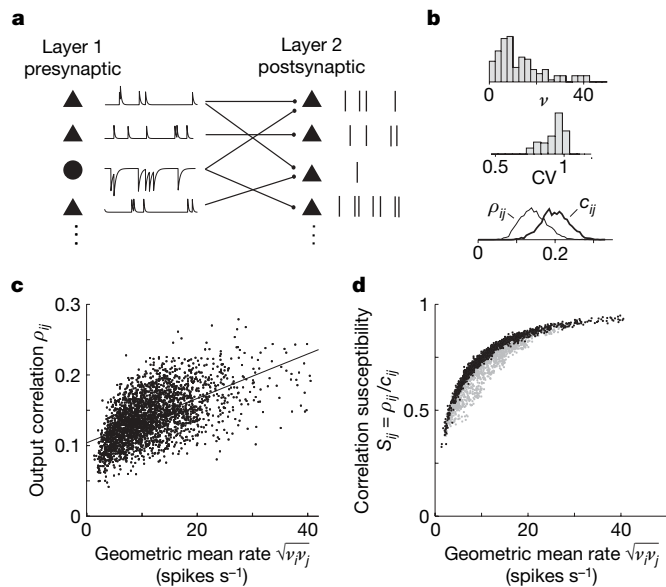


Figure 3 | Correlation-rate relationship in a simple network. **a**, Two-layer network with sparse random connectivity (connection probability was 0.25) where layer 1 contained excitatory and inhibitory cells firing with Poisson statistics. Layer 2 cells were conductance-based LIF neurons with randomly chosen synaptic weights. **b**, Distribution of v (top), CV (middle) and c_{ij} and ρ_{ij} (bottom) across layer 2. **c**, ρ_{ij} versus $\sqrt{v_i v_j}$ for layer 2 (linear regression $r = 0.57$). **d**, $S_{ij} = \rho_{ij}/c_{ij}$ versus $\sqrt{v_i v_j}$ for each pair shown in **c**. Pairs were grouped according to $q_{ij} = v_i/v_j$ (with $v_i \leq v_j$): $q_{ij} > 0.5$ (black) and $q_{ij} \leq 0.5$ (grey).

of excitatory or inhibitory neurons that fired stationary Poisson spike trains with fixed rate, whereas layer 2 neurons were conductance-based LIF model neurons that received randomly weighted AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) and GABA_A (γ -aminobutyric acid) synapses (see Methods). Each cell i in layer 2 received an input current with a different mean and variance, producing a natural heterogeneity in their firing rates, v_i , and variability, CV _{i} (Fig. 3b). In addition, each pair of layer 2 cells, i and j , shared a different number of afferents, giving rise to a wide distribution of input correlation coefficients, c_{ij} , and in turn a wide distribution of output correlation coefficients, ρ_{ij} (Fig. 3b).

A plot of ρ_{ij} versus the geometric mean of the rates $\sqrt{v_i v_j}$ shows a positive correlation (Fig. 3c). However, the trend is not nearly as tight as the one obtained in the controlled setting (compare Fig. 3c with Fig. 2c), primarily owing to the heterogeneity in c_{ij} (Fig. 3b, bottom panel). After normalization by the measured c_{ij} values (obtainable experimentally using intracellular recordings), we obtained $S_{ij} = \rho_{ij}/c_{ij}$ and a tight clustering of points along the same stereotyped curve emerged (Fig. 3d). If we group the pairs according to how different their rates are, we find that spike trains with similar rates (Fig. 3d, black points) showed slightly larger S_{ij} and a tighter correlation-rate relationship than pairs with more different rates (Fig. 3d, grey points). The same qualitative results were obtained in networks where synapses had slow kinetics (NMDA (N-methyl-D-aspartate) and GABA_B), or where the total conductance was highly non-gaussian (see Supplementary Information).

The experiments, simulations and theory, although showing a clear relationship between ρ and v , do not give a mechanistic understanding of the relationship. To develop this intuition, we examined phenomenological neuron models defined by a simple function, f , that transforms an input random variable, I , with gaussian statistics (μ, σ) to an output random variable, n (Fig. 4a). The inputs I_1 and I_2 to a pair of such neuron models were constructed in analogy to our two-cell system so that their correlation coefficient was c (compare Fig. 1a to Fig. 4a). We varied the mean input, μ , while keeping σ and c fixed (as in Fig. 2c, red curve), and computed the mean output $\langle n \rangle$ and the output correlation coefficient ρ (equation (2)). When f was strictly linear in I (Fig. 4c, top row), then $\rho = c$ for all values of μ , making ρ independent of $\langle n \rangle$. However, when f was threshold-linear, which captures the rectifying property of spike generation, ρ increased with $\langle n \rangle$ (Fig. 4b), as was observed previously (Figs 1e and 2c). The fact that the simple threshold-linear model reproduced the correlation-rate relationship indicates that the specific dynamics of the cell model are not vital to the relationship. Modifying the threshold-linear transfer, to include saturation, produced a non-monotonic relationship between ρ and $\langle n \rangle$ (Fig. 4c, middle row). Similarly, adding a boosting non-linearity to the threshold-linear model introduces a small concavity in the ρ - $\langle n \rangle$ relationship (Fig. 4c, bottom row). Despite this diversity of behaviour across these examples, one general feature is clear: when the input distribution $P(I)$ (coloured gaussians) overlapped with the nonlinear region of f , then ρ was significantly less than c .

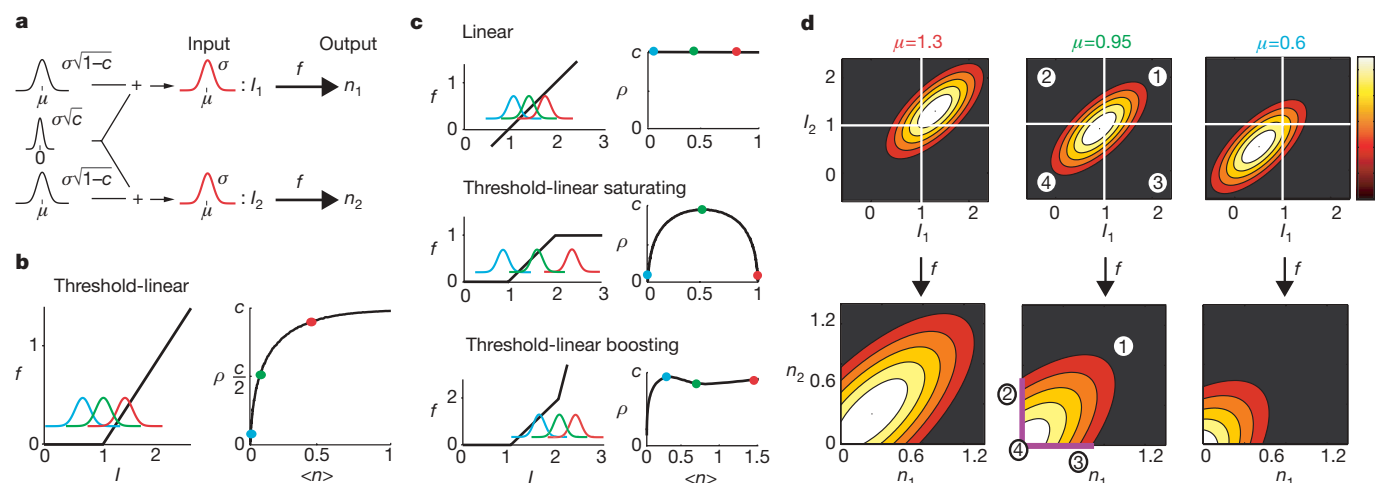


Figure 4 | Nonlinearities shape the correlation-rate relationship in a phenomenological neural model. **a**, The transfer $n_i = f(I_i)$ for a pair of models; note the parallels with Fig. 1a. **b**, f and the ρ - v relationship for the threshold-linear model. Overlaid are the input densities $P(I)$ (coloured gaussians) for three values of μ . **c**, Same as in **b** using three different f functions (see text). **d**, The joint densities $P(I_1, I_2)$ and $P(n_1, n_2)$ for the

threshold-linear model for the three values of μ in **b**. The threshold (white lines) partitions the input plane (I_1, I_2) into four quadrants. $P(I_1, I_2)$ is transformed differently in each quadrant (middle column). In **b-d** we set $\sigma = 0.3$ and the results are valid for c in the interval (0,0.5), whereas $c = 0.5$ in panel **c**.

For the threshold and saturating nonlinearities this effect was especially prominent as ρ limited to zero.

The correlation between I_1 and I_2 (or n_1 and n_2) is graphically represented as the eccentricity of their joint probability density function, $P(I_1, I_2)$ (or $P(n_1, n_2)$). The eccentricity, loosely defined as the elongation of the probability density function along the diagonal $I_1 = I_2$, increases with correlation. The eccentricity of $P(I_1, I_2)$ is, by construction, uniquely determined by c , so that for $c = 0$ the probability density function has a circular distribution, whereas as c approaches 1 the probability density function becomes increasingly confined to the diagonal. In contrast, the eccentricity of $P(n_1, n_2)$ depends on both c and μ (Fig. 4d, bottom row), as the following analysis using the threshold-linear model shows. The threshold, associated with f , partitions the input plane (I_1, I_2) into four quadrants (Fig. 4d, middle column). In each quadrant, the transformation of $P(I_1, I_2)$ into $P(n_1, n_2)$ via f is different. When μ is large, most of the mass of $P(I_1, I_2)$ is contained in quadrant 1, so that $P(I_1, I_2)$ is mapped approximately linearly to $P(n_1, n_2)$. As a consequence, $P(n_1, n_2)$ inherits much of the eccentricity of $P(I_1, I_2)$ (Fig. 4d, bottom-left panel), implying $\rho \approx c$. As μ decreases, the mass of $P(I_1, I_2)$ is shifted from quadrant 1 to the other three quadrants, and f distorts $P(I_1, I_2)$ so that the eccentricity of $P(n_1, n_2)$, and hence ρ , decreases (Fig. 4d, middle and right panels of bottom row). Thus, as μ is varied both ρ and $\langle n \rangle$ are expected to co-vary, yielding the correlation–rate relationship shown in Figs 1e, 2c and 4b. An analogous description can be made using the spiking LIF model (see Supplementary Information).

The above analysis shows that nonlinearities in the transfer function can cause different v -dependencies of ρ and allows us to explore the effect of other nonlinearities present in neurons. Phenomena such as spike refractoriness (see Fig. 6A of ref. 27), synaptic depression (see Fig. 2A of ref. 28), or dendritic sublinear summation (see Fig. 2 of ref. 29) impose a saturation on rate that would cause the ρ – v relationship to reach a maximum and begin to decrease (Fig. 4c, middle row). Dendritic supra-linear summation (Fig. 4A of ref. 30 and Fig. 2 of ref. 29) introduces boosting effects that would yield small inflections in the ρ – v relationship (Fig. 4c, bottom row). These observations predict that single neuron/synaptic nonlinearities can be an important determinant of network correlation.

A relationship between firing rate and pairwise correlation has, until now, been elusive^{8,10,17,18}. A consequence of the correlation–rate relationship is that, if the firing rate of a population of neurons is tuned to a certain feature (for example, orientation of a bar), the population correlation inherits the same tuning. Consistent with this prediction, ρ_T measured from neurons in primary visual cortex shows a tuning to stimulus orientation when $T < 50$ ms⁸. A direct test of the correlation–rate relationship would require intracellular recordings where measurements of both the afferent current and spike train correlations, as a stimulus is varied, could be obtained²⁴. Nevertheless, analysis of extracellular multi-unit recordings from macaque V1 (A. Kohn and M. Smith, personal communication) and electrosensory receptors (J. Middleton, J. Benda, A. Longtin and L. Maler, personal communication) shows a correlation–rate relationship similar to the one obtained in our network (Fig. 3c), supporting the generality of our result. The correlation–rate relationship prompts a re-examination of rate- and correlation-based coding hypotheses^{8,9,11–13,16,21,22}, because these measures are mechanistically intertwined.

METHODS SUMMARY

Whole-cell somatic recordings were made from layer 5 neurons in slices (300 μ m) from auditory and somatosensory cortices of young mice (postnatal day P14 to P21). Surgical, slicing and recording techniques were done as described previously¹⁹ and followed guidelines established by the NYU Animal Welfare Committee. Computer-generated currents (equation (1); duration $L = 1$ s) were delivered to 1–4 neurons every 1–5 s for a total of $N = 100$ trials. The terms $\xi_i(t)$ and $\xi_c(t)$ were gaussian noises low-pass-filtered with a time constant of 5 ms. The spike count from the i th cell in the k th trial, $n_i^k(t)$, was the number of spikes that occurred in the temporal window $(t, t + T)$. Time was

discretized with $\Delta t = 1.2$ ms. The spike count shift-corrected covariance was:

$$\text{Cov}(n_i, n_j) = \frac{1}{N(L-T)} \sum_{k=1}^N \sum_{t=0}^{L-T} (n_i^k(t)n_j^k(t) - n_i^k(t)n_j^{k+1}(t)) \quad (5)$$

where the interior sum ranges over time bins. The variance was $\text{Var}(n_i) = \text{Cov}(n_i, n_i)$ and the correlation coefficient ρ_T was computed as in equation (2). In Fig. 1 we used $T = 40$ ms (see Supplementary Information for analysis using $T = 10$ –300 ms).

We used a leaky integrate-and-fire (LIF) neural model²⁵ which obeyed: $\tau_m \frac{dV_i}{dt} = -V_i + \mu + \sigma \sqrt{\tau_m} (\sqrt{1 - c} \xi_i(t) + \sqrt{c} \xi_c(t))$ where V_i is the membrane potential of the i th neuron, and $\tau_m = 10$ ms, threshold = 20 mV and reset = 0 mV. Here $\xi_i(t)$ and $\xi_c(t)$ are white noise currents. Correlation coefficients ρ , defined as the limit of ρ_T as T increases, were estimated from the ratio of the area of the cross-correlation function normalized by the areas of the auto-correlation functions¹⁰ (see Supplementary Information). Other parameters that differ from the experimental data analysis were $L = 100$ s and $\Delta t = 0.5$ ms. The network in Fig. 3 had $N_E = 800$ excitatory and $N_I = 200$ inhibitory cells in layer 1 randomly connected (probability = 0.25) with 100 cells in layer 2. Layer 1 neurons fired Poisson trains with a rate of 10 spikes s^{-1} . Layer 2 neurons were LIF models with conductance-based excitatory and inhibitory synapses with time constants $\tau_E = 4$ ms, $\tau_I = 8$ ms, mean weights $J_E = 0.32$ nS, $J_I = 1.4$ nS and standard deviation $\Delta x = 0.35 J_x$ (where $x = E, I$).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 April; accepted 18 June 2007.

1. Mastronarde, D. N. Correlated firing of cat retinal ganglion cells. I. spontaneously active inputs to x- and y-cells. *J. Neurophysiol.* **49**, 303–324 (1983).
2. Schneidman, E., Berry, M. J., Segev, R. & Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012 (2006).
3. Shlens, J. et al. The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* **26**, 8254–8266 (2006).
4. Stopfer, M., Bhagavan, S., Smith, B. H. & Laurent, G. Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* **390**, 70–74 (1997).
5. Alonso, J. M., Usrey, W. M. & Reid, W. M. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature* **383**, 815–819 (1996).
6. Bruno, R. M. & Sakmann, B. Cortex is driven by weak but synchronously active thalamocortical synapses. *Science* **312**, 1622–1627 (2006).
7. Zohary, E., Shadlen, M. N. & Newsome, W. T. Correlated neuronal discharge rate and its implication for psychophysical performance. *Nature* **370**, 140–143 (1994).
8. Kohn, A. & Smith, M. A. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* **25**, 3661–3673 (2005).
9. deCharms, R. C. & Merzenich, M. M. Primary cortical representation of sounds by the coordination of action potentials. *Nature* **381**, 610–613 (1996).
10. Bair, W., Zohary, E. & Newsome, W. T. Correlated firing in macaque visual area MT: Time scales and relationship to behavior. *J. Neurosci.* **21**, 1676–1697 (2001).
11. Steinmetz, P. N. et al. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* **404**, 187–190 (2000).
12. Vaadia, E. et al. Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* **373**, 515–518 (1995).
13. Abeles, M. *Corticonics: Neural circuits of the cerebral cortex* (Cambridge Univ. Press, New York, 1991).
14. Svirkis, G. & Hounsgaard, J. Influence of membrane properties on spike synchronization in neurons: theory and experiments. *Network Comput. Neural Syst.* **14**, 747–763 (2003).
15. Galán, R. F., Fourcaud-Trocme, N., Ermentrout, G. B. & Urban, N. N. Correlation-induced synchronization of oscillations in olfactory bulb neurons. *J. Neurosci.* **26**, 3646–3655 (2006).
16. Shadlen, M. N. & Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
17. Binder, M. D. & Powers, R. K. Relationship between simulated common synaptic input and discharge synchrony in cat spinal motoneurons. *J. Neurophysiol.* **86**, 2266–2275 (2001).
18. Dorn, J. D. & Ringach, D. L. Estimating membrane voltage correlations from extracellular spike trains. *J. Neurophysiol.* **89**, 2271–2278 (2003).
19. Reyes, A. D. Synchrony-dependent propagation of firing rate in iteratively constructed networks *in vitro*. *Nature Neurosci.* **6**, 593–599 (2003).
20. Doiron, B., Rinzal, J. & Reyes, A. Stochastic synchronization in finite size spiking networks. *Phys. Rev. E* **74**, 030903 (2006).
21. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nature Rev. Neurosci.* **7**, 358–366 (2006).
22. Salinas, E. & Sejnowski, T. J. Correlated neuronal activity and the flow of neural information. *Nature Rev. Neurosci.* **2**, 539–550 (2001).
23. Destexhe, A., Rudolph, M. & Paré, D. The high-conductance state of neocortical neurons *in vivo*. *Nature Rev. Neurosci.* **4**, 739–751 (2003).

24. Lampl, I., Reichova, I. & Ferster, D. Synchronous membrane potential fluctuations in neurons of the cat visual cortex. *Neuron* **22**, 361–374 (1999).
25. Ricciardi, L. M. *Diffusion Processes and Related Topics in Biology* (Springer, Berlin, 1977).
26. Lindner, B., Doiron, B. & Longtin, A. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys. Rev. E* **72**, 061919 (2005).
27. Lau, D. et al. Impaired fast-spiking, suppressed cortical inhibition, and increased susceptibility to seizures in mice lacking kv3.2 K⁺ channel proteins. *J. Neurosci.* **20**, 9071–9085 (2000).
28. Tsodyks, M. V. & Markram, H. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl Acad. Sci. USA* **94**, 719–723 (1997).
29. Polsky, A., Mel, B. & Schiller, J. Computational subunits in thin dendrites of pyramidal cells. *Nature Neurosci.* **7**, 621–627 (2004).
30. Larkum, M. E., Senn, W. & Lüscher, H.-R. Top-down dendritic input increases the gain of layer 5 pyramidal neurons. *Cereb. Cortex* **14**, 1059–1070 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Colbert, A. Kohn, L. Maler, D. Nikolic, A.-M. Oswald and A. Renart for their critical reading of the manuscript, and R. Moreno-Bote, M. Schiff and J. Rinzel for insightful discussions. Funding was provided by the Spanish MEC (J.R.), HFSP (B.D.), a Burroughs Wellcome Fund career award and an NSF postdoctoral fellowship (E.S.-B.), NSF (K.J.) and NIH (A.R.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.D. (bdoiron@cns.nyu.edu) or J.R. (jrocha@cns.nyu.edu).

METHODS

Experiments. Surgical, slicing and recording techniques were done as described previously¹⁹ and followed guidelines established by the NYU Animal Welfare Committee. Slices were made from auditory and somatosensory cortices of young mice (postnatal day P14 to P21). Whole-cell somatic recordings were made from layer 5 neurons identified under an infrared video microscopy. During recordings, slices (300 μm thick) were perfused at room temperature or at 32 °C with ACSF (with 95% O₂, 5% CO₂). Pipettes were filled with 100 mM K-gluconate, 20 mM KCl, 10 mM phosphocreatine, 10 mM HEPES, 4 mM ATP-Mg, and 0.3 mM GTP at pH 7.3. Filled electrode resistances were in the range of 5 to 10 M Ω and recordings were performed under current-clamp conditions. Voltage and current signals were filtered at 10 KHz and digitized at 5 KHz.

Stimulus protocol. Computer-generated currents (equation (2); duration $L = 1$ s after removing the initial 100 ms) were delivered to 1–4 neurons every 1–5 s. The common $\xi_c(t)$ and the independent terms $\xi_i(t)$ were gaussian noises low-pass-filtered with a time constant of 5 ms. We presented each $N = 100$ trials stimulus block for 1–7 repetitions (see Supplementary Information for details).

Data analysis. Voltage traces were differentiated and action potential times were detected by setting a threshold on the differential traces. The spike trains emitted by the i th cell in the k th trial were represented by a binary time series, $y_i^k(t)$, which equalled 1 if there was a spike at time t and zero otherwise. The spike count, $n_i^k(t)$, was simply the number of spikes that occurred in a window of length T , that is, $n_i^k(t) = \sum_{t'=t}^{t+T} y_i^k(t')$. Time was discretized with $\Delta t = 1.2$ ms resolution and the first 100 ms of the spike trains were removed to correct for spike frequency adaptation. The spike count shift-corrected covariance was computed as:

$$\text{Cov}(n_i, n_j) = \frac{1}{N(L-T)} \sum_{k=1}^N \sum_{t=0}^{L-T} n_i^k(t) n_j^k(t) - \frac{1}{N(L-T)} \sum_{k=1}^N \sum_{t=0}^{L-T} n_i^k(t) n_j^{k+1}(t) \quad (6)$$

where the interior sums range over time bins. Note that in the second term, named the shift-corrector, spike counts from consecutive trials were multiplied. This term equals the product $\langle n_i \rangle \langle n_j \rangle$ for data that are stationary across trials (like in our LIF model) and corrected for both ‘chance correlations’ (baseline correlation expected for independent spike trains) and co-fluctuations due to drifts in the excitability across trials. The variance was obtained by noting that $\text{Var}(n_i) = \text{Cov}(n_i, n_i)$ and the correlation coefficient ρ_T was computed as in equation (2). The analysis shown in Fig. 1 used $T = 40$ ms whereas in the Supplementary Information we explore the range $T = 10$ –300 ms. The output spike shift-corrected cross-correlation function is defined as:

$$C_{ij}(\tau) = \frac{1}{N\Delta t(L-|\tau|)} \sum_{k=1}^N \sum_{t=0}^{L-|\tau|} y_i^k(t) y_j^k(t+\tau) - \frac{1}{N\Delta t(L-|\tau|)} \sum_{k=1}^N \sum_{t=0}^{L-|\tau|} y_i^k(t) y_j^{k+1}(t+\tau) \quad (7)$$

The auto-correlation function was obtained by making $i = j$ in equation (7). Note that, as in the covariance, we have corrected for ‘chance correlations’ in $C_{ij}(\tau)$ by subtracting the corresponding shift-corrector.

Computational model. We used a leaky integrate-and-fire (LIF) neural model²⁵ integrating white noise currents:

$$\tau_m \frac{dV_i}{dt} = -V_i + \mu + \sigma \sqrt{\tau_m} \left(\sqrt{1-c} \xi_i(t) + \sqrt{c} \xi_c(t) \right) \quad (8)$$

where V_i is the membrane potential of the i th neuron, and $\tau_m = 10$ ms, threshold = 20 mV and reset = 0 mV. Correlation coefficients ρ defined as the limit of ρ_T as T increases, were estimated by computing the ratio of the area of the cross-correlation function, $C_{ij}(\tau)$, normalized by the areas of the auto-correlation functions, $C_{ii}(\tau)$ ¹⁰ (see Supplementary Information). Other parameters in the simulations that differ from the data analysis were $L = 100$ s and $\Delta t = 0.5$ ms. ρ in Fig. 2c was numerically obtained from a linear regression of $\rho(c)$ in the interval $c = 0$ –0.12. Parameter values in Fig. 2: $\mu = 10, 14$ and 26 mV, $\sigma = 1.3$ mV (panel a); $\mu = 10$ mV, $\sigma = 1.3, 2.4$ and 8.8 mV (panel b); $\mu = 10$ –38 mV and $\sigma = 1.3$ mV (red in panel c); $\mu = 10$ mV and $\sigma = 1.3$ –15.8 mV (blue in panel c). The theoretical curves in Fig. 2 were computed from standard first passage time formulae appropriate for a LIF neuron receiving white noise inputs²⁵ (see Supplementary Information).

Network simulation. The network in Fig. 3 had $N_E = 800$ excitatory (E) and $N_I = 200$ inhibitory (I) cells in layer 1 and 100 cells in layer 2. Connections were randomly established from layer 1 to layer 2 with a probability $P = 0.25$. Layer 1 cells fired Poisson spike trains at rate $v_E = v_I = 10$ spikes s^{-1} . Each neuron in layer 2 was a LIF neuron model with conductance-based E and I synapses described by:

$$C_m \frac{dV}{dt} = g_L(E_L - V) + g_E(t)(E_E - V) + g_I(t)(E_I - V) \quad (9)$$

$$\frac{dg_\alpha}{dt} = -\frac{g_\alpha}{\tau_\alpha} + \sum_j J_{\alpha,j} \sum_k \delta(t - t_{\alpha,j}^k) \quad (\alpha = E, I) \quad (10)$$

where $\{t^k\}$ are the times of the spikes emitted by the j th pre-synaptic cell and $J_{\alpha,j}$ is the synaptic weight which was zero if the j th cell did not establish a connection or was randomly chosen from a gaussian distribution with mean $J_E = 0.32$ nS ($J_I = 1.4$ nS) and standard deviation $\Delta_\alpha = 0.35 J_\alpha$ ($\alpha = E, I$) if there was one. The synaptic time constants were $\tau_E = 4$ ms, $\tau_I = 8$ ms resembling AMPA and GABA_A kinetics (see Supplementary Information for a slow synapses analysis). We also set threshold = −55 mV, reset = −65 mV, refractory period = 2 ms, membrane capacitance $C_m = 0.12$ nF, leak conductance $g_L = 4.5$ nS, excitatory reversal potential $E_E = 0$ mV, and inhibitory reversal potential $E_I = -80$ mV.

LKB1 modulates lung cancer differentiation and metastasis

Hongbin Ji^{1,4*}, Matthew R. Ramsey^{10,12*}, D. Neil Hayes¹¹, Cheng Fan¹⁰, Kate McNamara^{1,4}, Piotr Kozlowski⁵, Chad Torrice¹¹, Michael C. Wu³, Takeshi Shimamura¹, Samantha A. Perera^{1,4}, Mei-Chih Liang^{1,4}, Dongpo Cai¹, George N. Naumov⁸, Lei Bao¹³, Cristina M. Contreras¹⁴, Danan Li^{1,4}, Liang Chen^{1,4}, Janakiraman Krishnamurthy^{10,11}, Jussi Koivunen¹, Lucian R. Chirieac⁶, Robert F. Padera⁶, Roderick T. Bronson⁹, Neal I. Lindeman⁶, David C. Christiani², Xihong Lin³, Geoffrey I. Shapiro^{1,7}, Pasi A. Jänne^{1,7}, Bruce E. Johnson^{1,7}, Matthew Meyerson^{1,15}, David J. Kwiatkowski¹⁵, Diego H. Castrillon¹⁴, Nabeel Bardeesy¹⁶, Norman E. Sharpless^{10,11,12} & Kwok-Kin Wong^{1,7}

Germline mutation in serine/threonine kinase 11 (*STK11*, also called *LKB1*) results in Peutz–Jeghers syndrome, characterized by intestinal hamartomas and increased incidence of epithelial cancers¹. Although uncommon in most sporadic cancers², inactivating somatic mutations of *LKB1* have been reported in primary human lung adenocarcinomas and derivative cell lines^{3–5}. Here we used a somatically activatable mutant *Kras*-driven model of mouse lung cancer to compare the role of *Lkb1* to other tumour suppressors in lung cancer. Although *Kras* mutation cooperated with loss of *p53* or *Ink4a/Arf* (also known as *Cdkn2a*) in this system, the strongest cooperation was seen with homozygous inactivation of *Lkb1*. *Lkb1*-deficient tumours demonstrated shorter latency, an expanded histological spectrum (adeno-, squamous and large-cell carcinoma) and more frequent metastasis compared to tumours lacking *p53* or *Ink4a/Arf*. Pulmonary tumorigenesis was also accelerated by hemizygous inactivation of *Lkb1*. Consistent with these findings, inactivation of *LKB1* was found in 34% and 19% of 144 analysed human lung adenocarcinomas and squamous cell carcinomas, respectively. Expression profiling in human lung cancer cell lines and mouse lung tumours identified a variety of metastasis-promoting genes, such as *NEDD9*, *VEGFC* and *CD24*, as targets of *LKB1* repression in lung cancer. These studies establish *LKB1* as a critical barrier to pulmonary tumorigenesis, controlling initiation, differentiation and metastasis.

Non-small-cell lung cancer (NSCLC) is subdivided into three subtypes: squamous cell carcinoma (SCC, ~28%), large-cell carcinoma (LCC, ~24%) and adenocarcinoma (~48%)⁶. Although these differ histologically⁷, each is highly lethal and, until recently, little clinical distinction was made among sub-types. Concomitant *KRAS* (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homologue) activation and *LKB1* inactivation occur in lung adenocarcinoma³, yet *Lkb1*^{-/-} mouse embryo fibroblasts are resistant to transformation by oncogenic *Ras* (ref.8), and mutations in *RAS* are rare in colonic polyps of Peutz–Jeghers syndrome (PJS) patients⁹. Given this tissue-specific relationship between *KRAS* activation and *LKB1* deficiency, we investigated the tumour suppressor function of *Lkb1* in the context of *Kras* activation in an *in vivo* mouse lung cancer model.

To discern relationships among pulmonary tumour suppressor mechanisms, we intercrossed a conditionally activatable *Lox-Stop-Lox Kras*^{G12D} (hereafter called *Kras*) allele¹⁰ and four conditional (L/L) or germline null (-/-) alleles: *Lkb1*^{L/L}, *p53*^{L/L}, *Ink4a/Arf*^{-/-} and *p16*^{Ink4a-/-} (refs 8, 11–13). Mice were inoculated with adenoviral CRE (*adeno-Cre*) by inhalation, transducing a small percentage of pulmonary cells¹⁰ (Supplementary Fig. 1). As previously described^{10,14–16}, mice with *Kras* activation developed tumours with high multiplicity, long latency and low aggressiveness (Table 1 and Supplementary Fig. 2). In contrast, concomitant *p16*^{Ink4a} and *p53* inactivation in animals lacking the *Kras* allele produced infrequent, highly lethal tumours. *Kras* activation cooperated potently with *p53* loss in lung tumorigenesis (Table 1, ref. 16), but only modestly with specific *p16*^{Ink4a} loss or combined *Ink4a/Arf* loss (Table 1); this indicates that the *Kras* mutation initiates tumorigenesis, whereas *p16*^{Ink4a} and *p53* constrain tumour progression.

Surprisingly, the strongest genetic interaction was *Kras* mutation combined with homozygous *Lkb1* inactivation (Table 1 and Supplementary Fig. 2). Although caution is warranted in comparing cohorts of mixed genetic backgrounds, median survival for *Kras Lkb1*^{L/L} or *Kras Lkb1*^{L/-} mice was 9 weeks after *adeno-Cre*-inoculation, compared with 14 weeks in *Kras p53*^{L/L} mice. Cooperation was also noted between *Kras* activation and heterozygous *Lkb1* inactivation, although loss of the wild-type allele was not seen in tumours of heterozygous mice (see below). Inactivation of *Lkb1* in the absence of *Kras* activation was insufficient for pulmonary neoplasia, establishing that *Lkb1* potently suppresses *Kras*-directed lung tumorigenesis *in vivo*.

Kras Lkb1^{L/-} or *Kras Lkb1*^{L/L} mice demonstrated significantly increased tumour burden at early time points (Fig. 1a and Supplementary Fig. 2) and larger (>3 mm) lesions at later time points than *Kras* mice (Fig. 1b). As previously reported^{10,15,16}, we did not detect metastasis or local invasion in *Kras* mice. Regional lymph-node metastases were observed in 7 of 22 *Kras Lkb1*^{L/-} or *Kras Lkb1*^{L/L} mice, and in 27 of 44 *Kras Lkb1*^{L/L} or *Kras Lkb1*^{L/-} mice, and axial skeleton metastases were found in 1 *Kras Lkb1*^{L/-} and in 4 *Kras Lkb1*^{L/L} or *Kras Lkb1*^{L/-} mice (Table 1 and Supplementary Fig. 2).

¹Department of Medical Oncology, Dana-Farber Cancer Institute, ²Department of Environmental Health, ³Department of Biostatistics, Harvard School of Public Health, ⁴Ludwig Center at Dana-Farber/Harvard Cancer Center, ⁵Division of Translational Medicine, ⁶Department of Pathology, ⁷Department of Medicine, Brigham and Women's Hospital, ⁸Department of Surgery, Children's Hospital, ⁹Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁰Department of Genetics, ¹¹Department of Medicine, ¹²Curriculum in Genetics and Molecular Biology, The Lineberger Comprehensive Cancer Center, The University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599, USA. ¹³Department of Molecular Sciences, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. ¹⁴Department of Pathology and Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9072, USA. ¹⁵Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ¹⁶Massachusetts General Hospital Cancer Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

*These authors contributed equally to this work.

Table 1 | Comparison of lung cancer cohorts

Genotype	Number treated	Median survival (weeks)*	Tumour multiplicity†	Squamous or mixed histology	Metastasis	Comments
<i>Kras</i>	26	24	Medium	0 of 16	0 of 19	See also refs 10, 14, 15
<i>Lkb1</i> ^{L/L} or <i>Lkb1</i> ^{L/L}	15	>40	NA	NA	NA	No tumours observed
<i>p53</i> ^{L/L}	16	>40	NA	0 of 1	NA	See also ref. 18
<i>p16</i> ^{Ink4a-/-} <i>p53</i> ^{L/L}	15	29	Low	0 of 5	NA	High frequency of fatal pulmonary haemorrhage
<i>Kras</i> <i>p16</i> ^{Ink4a-/-}	19	24	Medium	0 of 12	3 of 15 (20%)	
<i>Kras</i> <i>p53</i> ^{L/L}	17	14	High	0 of 9	4 of 9 (44%)	Compare with ref. 16; see also refs 14, 17
<i>Kras</i> <i>Ink4a/Arf</i> ^{-/-}	26	22	High	0 of 11	0 of 11	Compare with ref. 16
<i>Kras</i> <i>Lkb1</i> ^{L/+} or <i>Kras</i> <i>Lkb1</i> ^{L/+}	27	19	High	0 of 18	7 of 22 (32%)	
<i>Kras</i> <i>Lkb1</i> ^{L/L} or <i>Kras</i> <i>Lkb1</i> ^{L/L}	56	9	High	15 of 27 (56%)	27 of 44 (61%)	2 of 27 mice also demonstrated large-cell histology

NA, not applicable.

* Median latency shown is after *adeno-Cre* treatment at 5–6 weeks of age, estimated by Kaplan–Meier analysis.

† Tumour multiplicity: low, <3 per lung section; medium, 3–10 per lung section; high, >10 per lung section.

These results indicate that reduced *Lkb1* gene dosage facilitates progression and metastasis in *Kras*-induced lung cancers.

Consistent with previous reports^{10,14–18}, all tumours from *Kras* mice with or without *p16*^{Ink4a}, *Ink4a/Arf* or *p53* inactivation were

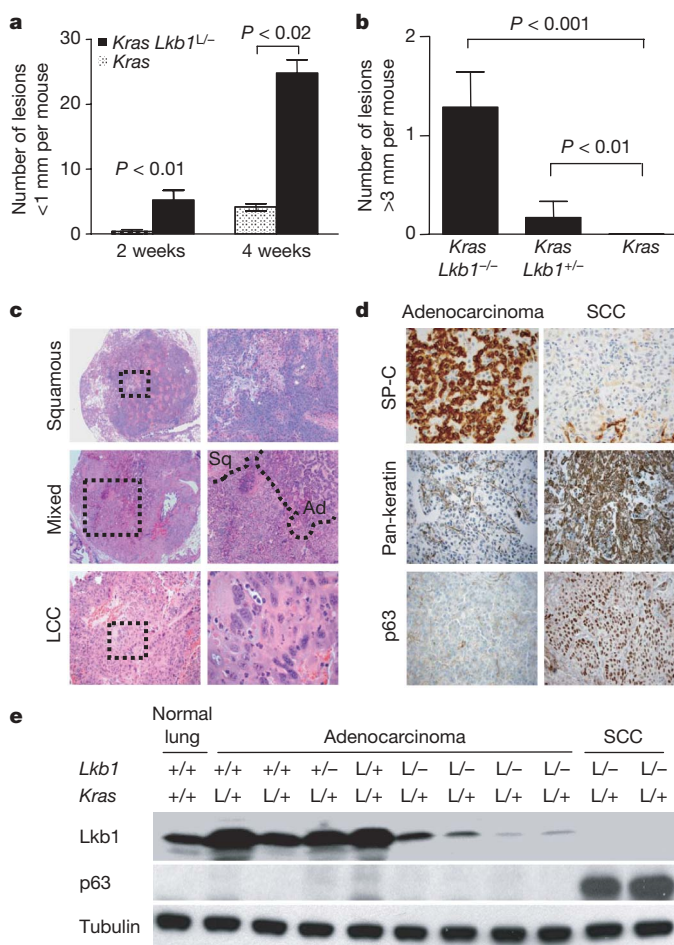


Figure 1 | Lung tumours in *Kras* and *Kras Lkb1*^{L/L} mice. **a**, Quantification of lesions (<1 mm) found in *Kras* or *Kras Lkb1*^{L/L} mice two and four weeks after *adeno-Cre* treatment. *n* = 4–6 for each group. Data are shown as mean ± s.e.m. **b**, Quantification of tumours (>3 mm) in *Kras Lkb1*^{L/L} mice 8 weeks after *adeno-Cre* treatment. Data are shown as mean ± s.e.m. **c**, Representative lung tumours from *Kras Lkb1*^{L/L} mice showing squamous (top), mixed (middle; Ad, adenocarcinoma; Sq, SCC) or large-cell (bottom) histology. The dotted box in the left image shows the area shown on the right. **d**, Immunohistochemical staining for SP-C, pan-keratin and p63 in adenocarcinomas (left) or in squamous tumours (right) from *Kras Lkb1*^{L/L} or *Kras Lkb1*^{L/L} mice. **e**, Western blot of *Lkb1* and p63 expression in tumours from mice of indicated genotype and histology. Histology is indicated as normal lung, adenocarcinoma or SCC. Tubulin serves as a loading control.

adenocarcinomas. In contrast, 15 of 27 *adeno-Cre*-treated *Kras Lkb1*^{L/L} or *Kras Lkb1*^{L/L} mice harboured SCCs or adenosquamous carcinomas whereas 2 of 27 lungs showed LCCs (Table 1, Fig. 1c). Squamous tumours from *Kras Lkb1*^{L/L} or *Kras Lkb1*^{L/L} mice did not express pro-surfactant protein C (SP-C), a marker of type II pneumocytes and adenocarcinomas, but expressed pan-keratin and p63, markers of SCC. SP-C expression was high and expression of pan-keratin and p63 was low or absent in adenocarcinomas (Fig. 1d, e). Therefore, *Lkb1* inactivation facilitated tumours of all three human histological subtypes.

To determine whether this mouse model recapitulated genetic events in human NSCLC, we assessed 144 human NSCLCs (Supplementary Table 1) for *KRAS* and *LKB1* mutations by direct exon sequencing, and investigated *LKB1* copy number status by multiplex ligation-dependent probe amplification¹⁹ (Supplementary Fig. 3). In accordance with other studies^{3–5}, 34% (27 of 80) adenocarcinomas harboured *LKB1* genomic alterations, predominantly (19 of 80) single-copy mutation or deletion (Table 2). *LKB1* alteration was seen in SCCs (8 of 42, 19%) with the predominant lesion (6 of 42) also being single-copy mutation or deletion. Additionally, *LKB1* alteration was found in 1 of 7 LCCs and 1 of 4 adenosquamous carcinomas. *KRAS* mutation has been reported in all subtypes of lung cancer² (Supplementary Table 2), and a subset of *LKB1* mutant adenocarcinoma and SCC also contained *KRAS* point mutations (Supplementary Tables 3–5). The frequent occurrence of single-copy mutation or deletion of *LKB1* in human tumours is consistent with the increased rate of tumour formation observed in *Kras Lkb1*^{L/+} or *Kras Lkb1*^{L/+} mice compared to *Kras* mice (Supplementary Fig. 2 and Table 2). Thus, inactivating mutations of *LKB1* are found in all histologic subtypes of human NSCLC.

The tumour-suppressor activity of *LKB1* has been reported to function through activation of *p53* and/or the *Ink4a/Arf* locus^{3,8,20}. Although we confirmed an effect of *Lkb1* loss on *p16*^{Ink4a} and *Arf* expression in mouse embryo fibroblasts (Supplementary Fig. 4), several lines of evidence indicate *Lkb1* also harbours *p53*- and *Ink4a/Arf*-independent tumour suppressor roles. Cooperation between *Lkb1* loss and *Kras* activation in mouse lung cancers was stronger than seen with loss of *p53* or *Ink4a/Arf* (Table 1). Moreover, *Lkb1*-deficient mouse tumours demonstrated enhanced metastasis and squamous differentiation—features not seen in the *p53*- or *Ink4a/Arf*-deficient backgrounds. A substantial fraction of human NSCLC harboured concomitant *p53* and *LKB1* mutation (Supplementary Tables 3–6), suggesting non-redundant roles in NSCLC. Lastly, *LKB1* reconstitution in human NSCLC cell lines lacking functional *p16*^{Ink4a}, *ARF* and *p53* demonstrated anti-tumour effects (see below). These data indicate that *LKB1* shows anti-tumour activity independent of *p16*^{Ink4a}, *ARF* and *p53*.

We performed unsupervised gene expression analysis of mouse primary lung tumours to elucidate previously unknown tumour-suppressor effects of *Lkb1*. Twenty-five tumours from 20 *Kras* mice of the indicated histologies and *Lkb1* genotypes were analysed using Affymetrix arrays (Supplementary Fig. 5, Supplementary Table 7 and

Table 2 | *LKB1* mutation incidence in human primary NSCLC samples

	Adenocarcinoma	SCC	Other*	Total
Only point mutation or deletion	8 of 80 (10)	4 of 42 (10)	1 of 22 (5)	13 of 144 (9)
Only single-copy loss by MLPA	11 of 80 (14)	2 of 42 (5)	1 of 22 (5)	14 of 144 (10)
Homozygous mutation or deletion	8† of 80 (10)	2 of 42 (5)	0 of 22 (0)	10 of 144 (7)
Total	27 of 80 (34)	8 of 42 (19)	2 of 22 (9)	37 of 144 (26)

Shown is the fraction of NSCLC samples studies found to harbour an *LKB1* mutation of the indicated class. Numbers in parentheses represent per cent.

* LCCs, adenosquamous carcinomas and NSCLCs of unknown subtype.

† Includes four samples with single-copy-loss by MLPA and point mutation by sequencing.

Supplementary Data 1). The most distinct group of tumours (tumours R–T) was comprised of squamous or adenosquamous (mixed) tumours from *Kras Lkb1*^{L/L} mice, which showed marked increases in the expression of genes (for example, p63, Krt5/6, desmoplakin; cluster A) overexpressed in human squamous lung cancer versus adenocarcinoma^{7,21}. These tumours also demonstrated sharply reduced expression of the *Lkb1* transcript and protein (Fig. 1e and data not shown). Therefore, in a subset of *Kras*-induced mouse tumours, loss of *Lkb1* expression was associated with the overexpression of transcripts characteristic of human SCCs.

Adenocarcinomas clustered into two groups on the basis of *Lkb1* status. Tumours A–L from *Kras Lkb1*^{+/+} mice showed high expression of *Lkb1* and transcripts associated with carbohydrate and lipid metabolism (Supplementary Fig. 5), consistent with the role of *Lkb1* in regulating the nutrient-sensing AMPK pathway²². These tumours showed increased phosphorylation of AMPK and ACC compared to adenocarcinoma tumours from *Kras Lkb1*^{L/+} or *Kras Lkb1*^{L/L} mice, which were characterized by reduced expression of *Lkb1* and reduced phospho-AMPK and phospho-ACC (Supplementary Fig. 6). These data indicate that *Lkb1* influences AMPK activation and expression of a large number of genes within the tumour subtype of Ras-driven adenocarcinoma.

Metastases from animals in all cohorts appeared adenocarcinoma-like, and squamous-like metastases were not seen. Accordingly, the *Lkb1*-deficient adenocarcinomas demonstrated increased expression of genes thought to promote angiogenesis and/or metastasis such as *Nedd9*, *Vegf*, *Loxl1*, *Pdgf* receptor and *Mmp2*, which were not increased in *Lkb1*-deficient squamous tumours (Supplementary Fig. 5). This indicates that *Lkb1* loss facilitates metastasis and permits squamous differentiation, but these effects appear genetically separable.

To understand the mechanism whereby *LKB1* suppresses metastasis, we stably expressed equivalent amounts of either wild-type *LKB1* or kinase-dead *LKB1*(K78I) (*LKB1*-KD) in *LKB1*-deficient NSCLC lines A549 or H2126 (Fig. 2a). Similar to a previous report on a breast cancer cell line²³, *LKB1* potently suppressed anchorage-independent growth in soft agar (Fig. 2b) and pulmonary metastases after tail-vein injection (Fig. 2c). Notably, *LKB1* reconstitution in the A549 line did not affect the expression of p53, p53 targets or the p53 response to ultraviolet radiation (Supplementary Fig. 7). Because this line harbours homozygous *INK4a/ARF* inactivation, this result demonstrates that the anti-metastatic effect of *LKB1* is *INK4a/ARF*- and *p53*-independent.

To refine the transcriptional analysis of *Lkb1*-deficient mouse tumours, A549 and H2126 cells with and without *LKB1* reconstitution were analysed by unsupervised expression profiling (Supplementary Fig. 8 and Supplementary Data 2). Only a small number of genes demonstrated reduced expression in both A549 and H2126 lines with *LKB1* reconstitution (Supplementary Fig. 8). Combined analysis of these human cell lines and *Lkb1*^{L/L} mouse tumours identified *NEDD9* as an *LKB1*-regulated transcript (Supplementary Figs 5 and 8), and a previous study has shown a role for *NEDD9* in metastasis²⁴. Protein expression of *NEDD9* in A549 and H2126 cell lines was repressed by wild-type *LKB1*, but not *LKB1*-KD (Fig. 2a), and increased *Nedd9* expression was noted in adenocarcinomas from *Kras Lkb1*^{L/L} or *Kras Lkb1*^{L/+} mice compared to adenocarcinomas from *Kras* mice or squamous tumours from *Kras Lkb1*^{L/L} or *Kras*

Lkb1^{L/L} (Fig. 2d). In addition, 70–80% reduction in *NEDD9* levels by short hairpin RNA (shRNA) in A549 cells reduced migration and invasion by more than 60% in Boyden chamber assays (Fig. 2e), suggesting that *NEDD9* is an important mediator of metastasis in mouse and human NSCLCs lacking *LKB1*.

Although *LKB1* negatively regulates the mTOR pathway²⁵ (Fig. 2a and Supplementary Fig. 6), treatment of parental A549 cells with rapamycin, an mTOR inhibitor, had no effect on *NEDD9* protein levels, despite a clear decrease in both S6 kinase and S6 ribosomal protein phosphorylation (Supplementary Fig. 9). This indicates that either *NEDD9* is a very stable protein (*T*_{1/2} > 72 h) or it is repressed by *LKB1* in an mTOR-independent manner.

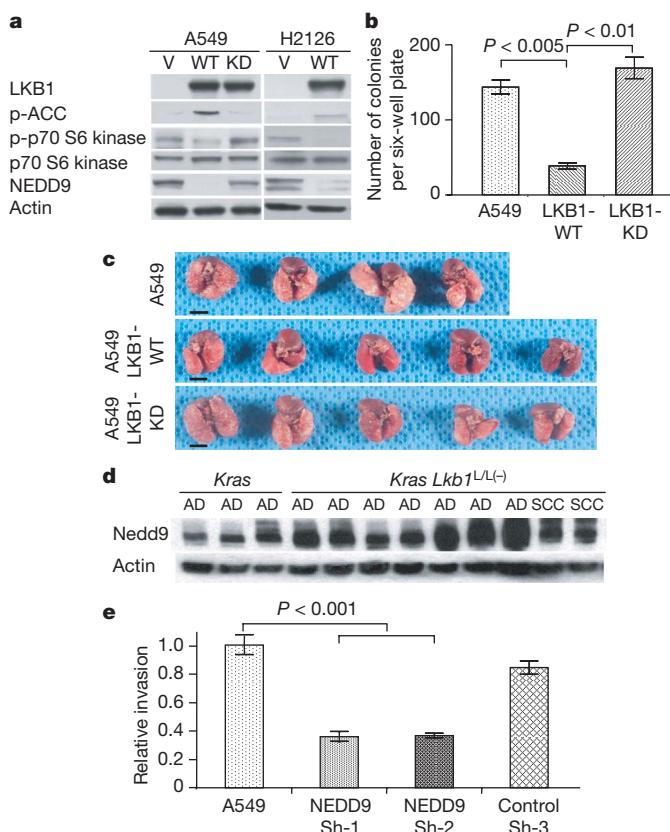


Figure 2 | *LKB1* suppresses metastasis. **a**, A549 and H2126 cells were stably transduced with pBABE-puro (V), pBABE-*LKB1* (WT) or kinase-dead pBABE-*LKB1*(K78I) (KD) and were assessed for expression of indicated genes by western blotting. Actin serves as a loading control. **b**, A549 cell lines were assessed for the ability to form colonies in soft agar. Data are shown as average number of colonies per six-well plate after 14 days (*n* = 4) and error bars represent s.e.m. **c**, A549 cell lines were injected into SCID mice by way of tail vein, and lung seeding was analysed eight weeks post-injection. Representative photographs of lungs are shown. **d**, Lung tumours from *Kras* or *Kras Lkb1*^{L/L} or *Kras Lkb1*^{L/+} mice were assessed for expression of *Nedd9*. Actin serves as a loading control. Tumour histology is indicated as adenocarcinoma (AD), adeno-squamous mixed (AS) or SCC. **e**, A549 parental cells and A549 stably transduced cells with two different shRNAs against *NEDD9* (Sh-1 and Sh-2) or control shRNA (Sh-3) were subjected to analyses with CytoSelect Cell Invasion assay. Invasion is plotted relative to A549 parental cell lines. Data are shown as mean ± s.e.m. (*n* = 6).

The LKB1 reconstitution experiments suggest that LKB1 modulates differentiation in NSCLC cell lines. In H2126 cells, LKB1 reproducibly altered the expression of hundreds of genes in independent experiments (Supplementary Fig. 8 and Supplementary Data 2). In particular, LKB1 restoration repressed characteristic luminal squamous epithelial markers such as keratins 8 and 18 and desmoplakin, and increased expression of adenomatous transcripts such as surfactants A, A2 and B (Supplementary Fig. 10). Intriguingly, LKB1 also increased expression of 'basal' markers such as Trp63 (also called p63 and p73L) and keratins 5, 6 and 17, and reduced expression of CD24—a luminal marker differentially regulated in putative breast and pancreatic carcinoma stem cells^{26–28}. CD24 was expressed in *Lkb1*-deficient squamous mouse tumours and in some adenocarcinomas (tumours O, P and R–T; Supplementary Fig. 5), and was repressed by LKB1 in A549 and H2126 cell lines (Supplementary Fig. 11). Given these results, it is tempting to speculate that LKB1 inactivation influences lineage choice in a common lung cancer progenitor cell.

We have described a mouse model of lung SCCs and have demonstrated repressive functions of LKB1 on metastasis through regulation of NEDD9. We find that LKB1 inactivation is frequent in all human NSCLC subtypes, suggesting similar tumour-suppressor roles in humans. Furthermore, our mouse data predict that LKB1 loss will serve as a prognostic marker of adverse disease. In accordance with this view, in ref. 29 it was recently noted that *LKB1* mutation correlates with advanced disease in NSCLCs, including SCCs. These data establish that LKB1 suppresses lung tumorigenesis through at least three independent mechanisms influencing tumour initiation, differentiation and metastasis.

METHODS SUMMARY

Kras mice were provided by T. Jacks, and were treated with 5×10^6 p.f.u. *adeno-Cre* (University of Iowa adenoviral core) intranasally as previously described¹⁰. For human tumour analysis, surgically resected human NSCLCs were flash-frozen, and genomic DNA was prepared from frozen tumour samples and sent for direct exon sequencing by Polymorphic DNA Technologies Inc. to detect *LKB1*, *KRAS*, *p53* and *p16^{INK4a}* mutations. Multiplex ligation-dependent probe amplification (MLPA) was performed as described previously¹⁹. Histology immunohistochemical and immunoblotting analyses were performed according to standard protocols, as previously described³⁰. Cytoselect cell invasion assays were performed as instructed by the manufacturer (Cell Biolabs). A549 cells (*INK4a/ARF*^{−/−}, *LKB1*^{−/−}) and H2126 cells (*INK4a/ARF*^{−/−}, *p53*^{−/−}, *LKB1*^{−/−}) were obtained from ATCC, and soft agar colony formation was assessed as described in Supplementary Data. For lung metastasis analysis, A549 cells or A549 stable cell lines with expression of wild-type LKB1 and LKB1-KD were injected into severe combined immunodeficient (SCID) mice by way of tail veins, and were assessed eight weeks post-inoculation for both gross inspection and histology analysis. For microarray analysis of human and mouse tumours, total RNA was extracted, amplified and labelled by standard methods and hybridized to Agilent 44,000 feature custom-designed Agilent arrays (human tumours) or to Mouse430A2 GeneChip Arrays (Affymetrix) representing 22,690 unique transcripts.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 March; accepted 19 June 2007.

Published online 5 August 2007.

- Hearle, N. *et al.* Frequency and spectrum of cancers in the Peutz–Jeghers syndrome. *Clin. Cancer Res.* **12**, 3209–3215 (2006).
- Forbes, S. *et al.* Cosmic 2005. *Br. J. Cancer* **94**, 318–322 (2006).
- Sanchez-Cespedes, M. *et al.* Inactivation of *LKB1/STK11* is a common event in adenocarcinomas of the lung. *Cancer Res.* **62**, 3659–3662 (2002).
- Carretero, J., Medina, P. P., Pio, R., Montuenga, L. M. & Sanchez-Cespedes, M. Novel and natural knockout lung cancer cell lines for the *LKB1/STK11* tumor suppressor gene. *Oncogene* **23**, 4037–4040 (2004).
- Avizienyte, E. *et al.* *LKB1* somatic mutations in sporadic tumors. *Am. J. Pathol.* **154**, 677–681 (1999).
- Tuveson, D. A. & Jacks, T. Modeling human lung cancer in mice: similarities and shortcomings. *Oncogene* **18**, 5318–5324 (1999).
- Hayes, D. N. *et al.* Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J. Clin. Oncol.* **24**, 5079–5090 (2006).
- Bardeesy, N. *et al.* Loss of the *Lkb1* tumour suppressor provokes intestinal polyposis but resistance to transformation. *Nature* **419**, 162–167 (2002).
- Entius, M. M. *et al.* Peutz–Jeghers polyps, dysplasia, and *K-ras* codon 12 mutations. *Gut* **41**, 320–322 (1997).
- Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic *K-ras*. *Genes Dev.* **15**, 3243–3248 (2001).
- Jonkers, J. *et al.* Synergistic tumor suppressor activity of *BRCA2* and *p53* in a conditional mouse model for breast cancer. *Nature Genet.* **29**, 418–425 (2001).
- Serrano, M. *et al.* Role of the *INK4a* locus in tumor suppression and cell mortality. *Cell* **85**, 27–37 (1996).
- Sharpless, N. E. *et al.* Loss of *p16^{INK4a}* with retention of *p19^{Arf}* predisposes mice to tumorigenesis. *Nature* **413**, 86–91 (2001).
- Johnson, L. *et al.* Somatic activation of the *K-ras* oncogene causes early onset lung cancer in mice. *Nature* **410**, 1111–1116 (2001).
- Meuwissen, R., Linn, S. C., van der Valk, M., Mooi, W. J. & Berns, A. Mouse model for lung tumorigenesis through Cre/lox controlled sporadic activation of the *K-Ras* oncogene. *Oncogene* **20**, 6551–6558 (2001).
- Fisher, G. H. *et al.* Induction and apoptotic regression of lung adenocarcinomas by regulation of a *K-Ras* transgene in the presence and absence of tumor suppressor genes. *Genes Dev.* **15**, 3249–3262 (2001).
- Jackson, E. L. *et al.* The differential effects of mutant *p53* alleles on advanced murine lung cancer. *Cancer Res.* **65**, 10280–10288 (2005).
- Meuwissen, R. *et al.* Induction of small cell lung cancer by somatic inactivation of both *Trp53* and *Rb1* in a conditional mouse model. *Cancer Cell* **4**, 181–189 (2003).
- Volikos, E. *et al.* *LKB1* exonic and whole gene deletions are a common cause of Peutz–Jeghers syndrome. *J. Med. Genet.* **43**, e18 (2006).
- Karuman, P. *et al.* The Peutz–Jegher gene product *LKB1* is a mediator of *p53*-dependent cell death. *Mol. Cell* **7**, 1307–1319 (2001).
- Raponi, M. *et al.* Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res.* **66**, 7466–7472 (2006).
- Hardie, D. G. New roles for the *LKB1*→AMPK pathway. *Curr. Opin. Cell Biol.* **17**, 167–173 (2005).
- Zhuang, Z. G., Di, G. H., Shen, Z. Z., Ding, J. & Shao, Z. M. Enhanced expression of *LKB1* in breast cancer cells attenuates angiogenesis, invasion, and metastatic potential. *Mol. Cancer Res.* **4**, 843–849 (2006).
- Kim, M. *et al.* Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene. *Cell* **125**, 1269–1281 (2006).
- Shaw, R. J. *et al.* The *LKB1* tumor suppressor negatively regulates mTOR signaling. *Cancer Cell* **6**, 91–99 (2004).
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100**, 3983–3988 (2003).
- Li, C. *et al.* Identification of pancreatic cancer stem cells. *Cancer Res.* **67**, 1030–1037 (2007).
- Shipitsin, M. *et al.* Molecular definition of breast tumor heterogeneity. *Cancer Cell* **11**, 259–273 (2007).
- Matsumoto, S. *et al.* Prevalence and specificity of *LKB1* genetic alterations in lung cancers. *Oncogene* advance online publication, doi: 10.1038/sj.onc.1210418 (26 March 2007).
- Ji, H. *et al.* *K-ras* activation generates an inflammatory response in lung tumors. *Oncogene* **25**, 2105–2112 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Tonon, C. Perou, W. Kim and the Harvard Lung SPOR group for advice and discussions; J. Yokota for sharing unpublished data; and W. Winckler, R. Mukundhan, S. Zaghlul, H. Xia, B. L. Jung, M. Zheng and C. Lam for technical support. We acknowledge the technical assistance of the UNC Tissue Procurement Core Facility, an NCI-designated core laboratory. This work was supported by the NIH (NIA and NCI), the Sidney Kimmel Foundation for Cancer Research (D.H.C., K.-K.W. and N.E.S.), the American Federation of Aging (N.E.S.), the Joan Scarangelo Foundation to Conquer Lung Cancer (K.-K.W.), the Flight Attendant Medical Research Institute (K.-K.W.), the Waxman Foundation (N.B.), the Harvard Stem Cell Institute (N.B.), and the Linda Verville Foundation (N.B.).

Author Contributions The laboratories of N.B., N.E.S. and K.-K.W. contributed equally to this work.

Author Information The entire set of unprocessed raw microarray data for both the human and the mouse analyses is available at <http://genome.unc.edu> and through the Gene Expression Omnibus (GSE6135). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.-K.W. (kwong1@partners.org) or N.E.S. (nes@med.unc.edu) or N.B. (nelbardeesy@partners.org).

METHODS

Mouse colony and mouse tumour analysis. All mice were housed and treated in accordance with protocols approved by the institutional care and use committees for animal research at the Dana-Farber Cancer Institute and the University of North Carolina. All cohorts in Table 1 were of a similar, mixed genetic background (~75% C57Bl/6, ~25% FVB/n and 129SvEv). More than 500 mice were analysed in a standard manner for experiments reported in Table 1, but *adeno-Cre*-treated littermates of less informative genotypes (for example, compound heterozygotes, etc.) and animals treated with empty adenovirus are not shown in the interest of brevity. In all cases, heterozygote mice showed tumour-prone phenotypes intermediate to the wild-type and homozygous mutant animals. For CRE expression, 5×10^6 p.f.u. *adeno-Cre* (purchased from University of Iowa adenoviral core) was administered intranasally as previously described¹⁰.

Human tumour analysis. All human studies were approved by the University of North Carolina and Massachusetts General Hospital institutional review boards. Surgically resected human NSCLCs of all clinical stages were flash-frozen and stored at -80°C until the time of analysis. Samples at University of North Carolina were obtained by means of the Tissue Procurement Facility. Genomic DNA was prepared from frozen tumour samples using Qiagen genomic DNA purification columns and sent for direct exon sequencing by Polymorphic DNA Technologies Inc. to detect *LKB1*, *KRAS*, *p53* and *p16^{INK4a}* mutations. MLPA was performed as described previously^{19,31} using the SALSA MLPA kit P101 STK11 (MRC-Holland), which includes 1 probe set for each of the 10 exons of *LKB1*, 5 probe sets extending 0.9 Mb 5' to *LKB1*, 1 probe set 10 Mb 3' to *LKB1*, and 12 control probe sets from elsewhere in the genome. Briefly, 3.4 μl genomic DNA (20 ng μl^{-1}) was incubated at 98°C for 5 min. After cooling to 21°C , 1 μl probe mix (containing 1 fmol of probes) and 1 μl SALSA hybridization buffer were added, and the solution was denatured at 95°C for 2 min and hybridized at 60°C for 16 h. Hybridized probes were ligated at 54°C for 15 min by addition of 21 μl ligation mixture. Following heat inactivation, 7.5 μl ligation reaction was mixed with 22.5 μl PCR buffer, heated to 60°C , mixed with 7.5 μl PCR mixture (SALSA polymerase, dNTPs and universal primers, one of which was labelled with fluorescein) and subjected to PCR amplification for 30 cycles.

Amplification products were diluted in water and then 1:9 in HiDi formamide (ABI) containing 1/36 volume of ROX500 size standard (ABI; final dilution 20-fold), and were then separated by size on an ABI 3100 Genetic Analyser (ABI). Electropherograms were analysed by GeneMapper v3.5 (ABI), and peak height data were exported to an Excel table. Excel programs were used to transform the peak height data to normalized values, such that control samples gave a value of 1 after normalization. Briefly, peak heights for each probe were divided by the average signal from five control probes (on chromosomes 4, 12, 15 and 16), and then that value was divided by a similar value calculated from reference samples. We used the average values from four reference blood DNA samples processed concurrently for each analysis.

Histology and immunohistochemistry. Mice were killed and the left lungs were dissected. The right lung and mediastinal structures were inflated with neutral buffered 10% formalin for 10 min and fixed in 10% formalin overnight at room temperature. Fixed tissues were embedded in paraffin, sectioned at 5 μm , and haematoxylin and eosin stained (Department of Pathology in Brigham and Women's Hospital). Immunohistochemical analyses were performed as described³⁰. The antibodies used were: SPC (AB3786, Chemicon), pan-keratin (Z0622, Dako), p63 (ab3239, Abcam), p-AMPK (2535, Cell Signaling), phospho-acetyl-CoA carboxylase (Ser79) (3661, Cell Signaling) and VEGFC (2712, Cell Signaling).

Western blotting and mRNA analysis. Western blot assays were performed as previously described³² with antibodies against *p16^{INK4a}* (M-156, Santa Cruz), Arf (ab-80, Abcam), actin (C-1, Santa Cruz), Lkb1 (Clone 5c10, Upstate), Nedd9 (gift of L. Chin), S6 kinase (9202, Cell Signaling), phospho-S6 kinase (9204, Cell Signaling), S6 ribosomal protein (2217, Cell Signaling), phospho-S6 ribosomal protein (2215, Cell Signaling), p21^{CIP1} (F-8, Santa Cruz), phospho-p53 (Cell Signaling), p53 (CM5, Novacastra), phospho-ACC (Cell Signaling), tubulin (clone DM 1A, Sigma-Aldrich), p63 (4892, Cell Signaling) and β -actin (A5441, Sigma). Expression of messenger RNA was analysed by quantitative TaqMan real-time PCR as previously described with some modifications³³. Reactions were carried out using complementary DNA equivalent to 80 ng

RNA and were performed in triplicate for each sample. 18S ribosomal RNA was used as a loading control for all reactions. The primer set for 18S (Hs99999901_s1) was purchased from Applied Biosystems; *p16^{INK4a}* and Arf primers were generated as previously described³³.

In vitro analysis. A549 cells (*INK4a/ARF^{-/-}*, *LKB1^{-/-}*) and H2126 cells (*INK4a/ARF^{-/-}*, *p53^{-/-}*, *LKB1^{-/-}*) were obtained from ATCC. To assess soft agar colony formation, parental A549 cells and A549 stable cell lines with expression of wild-type *LKB1* and *LKB1-KD* were suspended in a top layer of RPMI1640 containing 10% FBS and 0.4% Select agar (Gibco/Invitrogen) at 5,000 cells per well in triplicate in 6-well plates and plated on a bottom layer of RPMI1640 containing 10% FBS and 1% Select agar. After 2-week culture, cells were stained with 0.5 ml crystal violet for 1 h. The colonies were then counted in triplicate wells from 10 fields photographed with a $\times 10$ objective. Lentiviral-based shRNA constructs against human NEDD9 (sh-1, CCTCCTTCATACCACTCAA; sh-2, CCAGCAGAAACCACTGAGAAA) were provided by M. Kim and L. Chin.

In vivo lung seeding assay of NSCLC cell lines. Parental A549 cells and A549 stable cell lines with expression of wild-type *LKB1* and *LKB1-KD* were injected into SCID mice intravenously by way of tail veins. After eight weeks of inoculation, the mice were sacrificed and the lungs were dissected for both gross inspection and histology analysis.

Statistical analysis. Tumour-free survival and comparisons of tumour numbers and colonies in soft agar were analysed using Graphpad Prism4. Statistical analyses were performed using nonparametric Mann-Whitney test. Comparisons of mRNA levels were made using the unpaired Student's *t*-test. All data are shown as mean \pm s.e.m.

Microarray analysis. For mouse tumours, total RNA was extracted, amplified and labelled by standard methods and hybridized to Mouse430A2 GeneChip Arrays (Affymetrix) representing 22,690 unique transcripts. Probe-level intensity data files in the CEL format (<http://www.stat.lsa.umich.edu/~kshedden/Courses/Stat545/Notes/AffxFileFormats/cel.html>) were pre-processed using the Robust Multichip Average³⁴⁻³⁶ as implemented in Bioconductor (<http://www.bioconductor.org/>). Gene-expression data were filtered using low stringency, pre-defined criteria: probe set intensity (>32 in all samples) and dynamic variation (more than twofold over the entire sample set). After filtering, multiple probe sets that presented the same genes were collapsed by taking the median value for that gene per array yielding 3,275 unique genes, on which two-way hierarchical clustering was performed (Supplementary Fig. 5). Excerpted clusters are shown in Supplementary Fig. 5, and the list and normalized expression of all 3,275 filtered genes is available as Supplementary Data 1. For human cell lines, total RNA was extracted, amplified and labelled by standard methods and hybridized to Agilent 44,000 feature custom-designed Agilent arrays, which are largely based on the 44,000 feature Agilent Human catalogue arrays, and scanned using an Agilent scanner. Gene expression data were filtered using the same criteria as for the mouse tumours. This filtering strategy yielded 9,644 non-unique transcripts which were then analysed by hierarchical clustering (Supplementary Figs 8 and 10), with excerpted clusters shown. The entire list and normalized expression of all 9,644 filtered transcripts is available as Supplementary Data 2. The entire set of unprocessed raw data for both the human and mouse analyses is available at <http://genome.unc.edu> and through the Gene Expression Omnibus (GSE6135).

31. Kozlowski, P. *et al.* Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. *Hum. Genet.* **121**, 389–400 (2007).
32. Sharpless, N. E., Ramsey, M. R., Balasubramanian, P., Castrillon, D. H. & DePinho, R. A. The differential impact of *p16(INK4a)* or *p19(ARF)* deficiency on cell growth and tumorigenesis. *Oncogene* **23**, 379–385 (2004).
33. Krishnamurthy, J. *et al.* *p16INK4a* induces an age-dependent decline in islet regenerative potential. *Nature* **443**, 453–457 (2006).
34. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
35. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
36. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).

Cdk1 is sufficient to drive the mammalian cell cycle

David Santamaría^{1*}, Cédric Barrière^{1,2*}†, Antonio Cerqueira¹, Sarah Hunt¹†, Claudine Tardy¹, Kathryn Newton³, Javier F. Cáceres³, Pierre Dubus², Marcos Malumbres¹ & Mariano Barbacid¹

Unicellular organisms such as yeasts require a single cyclin-dependent kinase, Cdk1, to drive cell division¹. In contrast, mammalian cells are thought to require the sequential activation of at least four different cyclin-dependent kinases, Cdk2, Cdk3, Cdk4 and Cdk6, to drive cells through interphase, as well as Cdk1 to proceed through mitosis². This model has been challenged by recent genetic evidence that mice survive in the absence of individual interphase Cdk3^{3–8}. Moreover, most mouse cell types proliferate in the absence of two or even three interphase Cdk3^{8–10}. Similar results have been obtained on ablation of some of the activating subunits of Cdk3, such as the D-type and E-type cyclins^{11–14}. Here we show that mouse embryos lacking all interphase Cdk3 (Cdk2, Cdk3, Cdk4 and Cdk6) undergo organogenesis and develop to midgestation. In these embryos, Cdk1 binds to all cyclins, resulting in the phosphorylation of the retinoblastoma protein pRb and the expression of genes that are regulated by E2F transcription factors. Mouse embryonic fibroblasts derived from these embryos proliferate *in vitro*, albeit with an extended cell cycle due to inefficient inactivation of Rb proteins. However, they become immortal on continuous passage. We also report that embryos fail to develop to the morula and blastocyst stages in the absence of Cdk1. These results indicate that Cdk1 is the only essential cell cycle Cdk. Moreover, they show that in the absence of interphase Cdk3, Cdk1 can execute all the events that are required to drive cell division.

When *Cdk4*^{+/-};*Cdk6*^{+/-};*Cdk2*^{+/-} mice are crossed, 25% of the conceptuses should be triple knockout (TKO) *Cdk4*^{-/-};*Cdk6*^{-/-};*Cdk2*^{-/-}, as *Cdk4* and *Cdk2* are genetically linked^{9,10}. We obtained TKO embryos at the expected mendelian ratios from such crosses (Supplementary Table 1). These embryos underwent morphogenesis and organogenesis and developed until embryonic day 12.5 (E12.5) (Fig. 1a). Cells proliferated in most tissues, as measured by staining for the proliferative marker Ki67 (Fig. 1a, Supplementary Fig. 1a). However, TKO embryos were not normal and began to die at E13.5, with all dying by E15.5. The livers of live TKO embryos at E12.5–E13.5 showed a three-fold reduction in cellularity, a decrease that is significantly higher than expected from their smaller size (25–40%). The levels of haematopoietic stem cells, granulocyte-macrophage progenitors and megakaryocyte-erythroid progenitors were consistent with the overall reduction in liver cellularity. Only common myeloid progenitors showed a greater (eightfold) reduction (Supplementary Fig. 1b). More importantly, livers from TKO embryos, unlike those from *Cdk4*^{-/-};*Cdk6*^{-/-} embryos, undergo substantial apoptosis at E13.5–E14.5, as determined by active caspase 3 immunoreactivity (Supplementary Fig. 1b). The hearts of TKO embryos showed thinner ventricular walls than those of wild-type controls owing to a decrease in the number of proliferating cardiomyocytes (data not shown). This phenotype was indistinguishable from that of *Cdk4*^{-/-};*Cdk2*^{-/-}

embryos, which could complete embryonic development¹⁰. Thus, although we cannot rule out the possibility that the heart phenotype contributed to embryonic death, the severity of the haematopoietic defects makes them more likely to account for the failure of TKO embryos to develop beyond midgestation.

The loss of interphase Cdk3 did not have a significant effect on the levels of expression of other cell-cycle regulators. The overall levels of Cdk1, Cdk7 and Cdk9 appeared to be unchanged (Fig. 1b). There were no significant variations in the expression of the cyclins tested (D1, D2, E1, A2 and B1). Among Cip/Kip inhibitors, only cyclin-dependent kinase inhibitor 1B (*Cdkn1b*, also known as p27^{Kip1}) showed decreased protein levels in TKO embryos (Fig. 1b), probably owing to increased protein instability, as the levels of *Cdkn1b* messenger RNA were not affected (Supplementary Fig. 2). pRb was phosphorylated in TKO embryos, including at those residues (Ser 608, Ser 780 and Ser 807/811) that were previously thought to be specific targets of interphase Cdk3¹⁵ (Fig. 1c).

To understand the molecular mechanisms that drive cell division in the absence of interphase Cdk3, we investigated whether Cdk1 could interact with G1/S cyclins. As shown in Fig. 1d, Cdk1 interacts with cyclin D1 and cyclin D2 in lysates derived from embryos lacking Cdk4. Likewise, Cdk1 binds to cyclin E in *Cdk2*^{-/-} and *Cdk4*^{-/-};*Cdk2*^{-/-} embryos, as previously reported^{10,16}. Moreover, cyclin D and E immunoprecipitates from TKO embryos could phosphorylate recombinant pRb (Fig. 1e), possibly owing to their ability to form complexes with Cdk1.

The proliferation of primary mouse embryonic fibroblasts (MEFs) in culture was partially compromised in the absence of interphase Cdk3. However, most TKO MEF cultures (11 of 12) became immortal upon continuous passage (Supplementary Fig. 3a, b). These TKO MEFs, as well as control cultures that expressed at least one interphase Cdk, maintained the expression of cyclin-dependent kinase inhibitor 2A (*Cdkn2a*, also known as p16^{INK4a}) and a functional P53 response (data not shown). Immortal TKO cultures, unlike those expressing at least one interphase Cdk, retained their extended doubling time. To understand the molecular basis of the slower cell cycle of TKO MEFs, we infected them with a retrovirus that expressed a fragment of the simian virus 40 (SV40) large T antigen (T₁₂₁) that is known to inactivate the three members of the Rb family¹⁷. Inactivation of all pocket proteins made the proliferation rates of TKO MEFs indistinguishable from those of wild-type cells (Supplementary Fig. 3c). These observations indicate that the slower cell cycles of TKO MEFs are a consequence of incomplete inactivation of Rb proteins. We obtained similar results from TKO MEFs derived from immortal *Cdk4*^{-/-};*Cdk6*^{-/-};*Cdk2*^{lox/lox} cells after infection with a retrovirus expressing the Cre recombinase (Supplementary Fig. 3d), ruling out the idea that the proliferation of TKO MEFs was a consequence of embryonic plasticity.

¹Molecular Oncology Programme, Centro Nacional de Investigaciones Oncológicas (CNIO), E-28029 Madrid, Spain. ²EA2406 University of Bordeaux 2, F-33076 Bordeaux, France.

³MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK. †Present addresses: Laboratoire de Morphogenèse et Signalisation Cellulaire, UMR 144/Institut Curie, 25 rue d'Ulm, 75248 Paris Cedex 05, France (C.B.); Fundació Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol, Carretera de Can Ruti, Camí de les Escoles, 08916 Badalona, Barcelona, Spain (S.H.).

*These authors contributed equally to this work.

TKO MEFs exited quiescence when stimulated with serum, albeit with significantly delayed (6–8 h) kinetics (Fig. 2a). However, most primary and immortal TKO MEFs entered the cell cycle (Fig. 2b) and executed the molecular steps that are known to be required to proceed into S phase, including pRb phosphorylation, E2F-dependent induction of cyclin A2 expression and degradation of the Cdk inhibitor Cdkn1b (Fig. 2c). In addition, Cdk1 formed complexes with cyclin D1 as early as 12 h after serum stimulation. These complexes were observed only at later time points (beyond 24 h) in control MEFs (Fig. 2d). Thus, it is tempting to speculate that Cdk1–cyclin D complexes might be functionally relevant during mitotic exit and/or early G1 phase, a period when other cyclins are being actively targeted by the anaphase-promoting complex/cyclosome¹⁸.

Next, we investigated whether Cdk1–cyclin complexes were responsible for allowing TKO cells to enter S phase. We infected TKO MEFs with lentiviral vectors expressing either control or short hairpin RNAs (shRNAs) specific for *Cdc2a*, the locus that encodes Cdk1. Knockdown of Cdk1 had no effect on the ability of cells expressing Cdk4 and Cdk2 to exit G0 and enter S phase (Fig. 2e). However, Cdk1 depletion completely abrogated S phase entry in TKO MEFs, an effect that correlated with the inhibition of pRb phosphorylation

(Fig. 2e). These observations indicate that, in the absence of interphase Cdks, Cdk1 can be activated by D-type and/or E-type cyclins to phosphorylate pRb and bring cells out of quiescence.

Finally, we investigated whether Cdk1 was essential for cell division. To this end, we generated mutant mice that were heterozygous for *Cdc2a* from two independent embryonic stem cells that carried β -geo cassette insertions within this locus^{19,20} (Fig. 3a). These integration events generate a Cdk1– β -geo fusion protein that contains only the 12 amino-terminal residues of Cdk1 (Fig. 3b). Heterozygous *Cdc2a*^{+/-mut1} cells express about 50% as much Cdk1 as is present in wild-type cells (Fig. 3c). Crosses between heterozygous *Cdc2a*^{+/-mut1} or *Cdc2a*^{+/-mut2} mice did not yield homozygous *Cdc2a* mutant animals nor midgestation embryos (E10.5–E13.5) (Fig. 3d). Likewise, we could not identify E2.5 (morula stage) or E1.5 (2–4-cell) *Cdc2a*^{mut1/mut1} or *Cdc2a*^{mut2/mut2} embryos. Analysis of 23 morulae that were allowed to grow in culture for 5–6 days using nested primers also failed to identify homozygous *Cdc2a*^{mut1/mut1} embryos (Fig. 3d). Moreover, the percentage of wild-type and *Cdc2a*^{+/-mut1} embryos were those expected if homozygous embryos were not viable. Finally, crosses between wild-type and *Cdc2a*^{+/-mut1} male and female mice yielded the expected percentage of *Cdc2a*^{+/-mut1} offspring (data not shown), indicating that

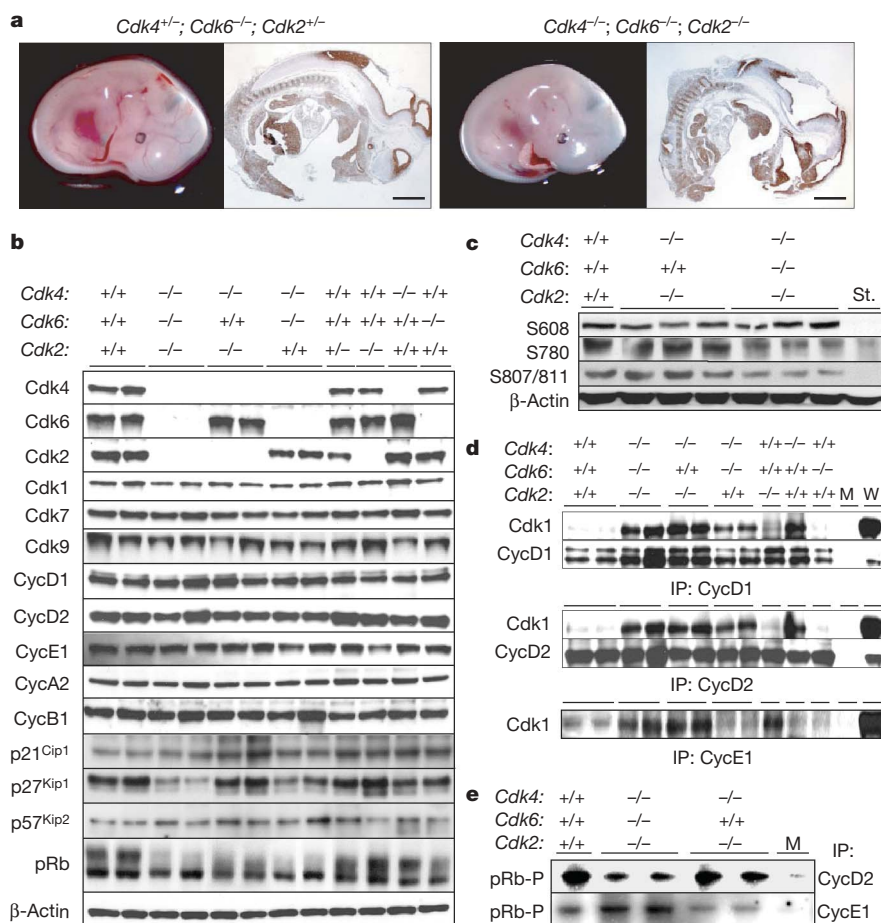


Figure 1 | Characterization of embryos lacking all interphase Cdks.

a, Embryos lacking all interphase Cdks undergo organogenesis and develop until midgestation. Left, *Cdk4*^{+/-};*Cdk6*^{-/-};*Cdk2*^{+/-} E12.5 embryos. Right, *Cdk4*^{-/-};*Cdk6*^{-/-};*Cdk2*^{-/-} E12.5 embryos. In both cases, the left image depicts a picture of the whole embryo and the right image depicts a Ki67-immunostained sagittal section (×12). Scale bar, 1 mm. **b**, Expression of cell cycle regulators in independent E12.5 embryos of the indicated genotype. Results from TKO embryos are shown next to results from wild-type embryos (left column). Protein extracts were analysed by immunoblotting with antibodies elicited against the indicated proteins (see Methods). Expression of β -actin served as loading control. **c**, Analysis of pRb phosphorylation in wild-type, *Cdk4*^{-/-};*Cdk6*^{+/-};*Cdk2*^{-/-} and

Cdk4^{-/-};*Cdk6*^{-/-};*Cdk2*^{-/-} E12.5 embryos using antibodies specific for phosphorylated residues S608, S780 and S807/811. St.: pRb phosphorylation at these residues in TKO serum-starved cells. **d**, Extracts from E12.5 embryos of the indicated genotype were immunoprecipitated with antibodies against cyclin D1 (top), cyclin D2 (middle) and cyclin E1 (bottom) and analysed by immunoblotting using antiserum against Cdk1, cyclin D1 and cyclin D2 as indicated. Results from two independent embryos are shown for wild-type, TKO and double mutant embryos. M, mock immunoprecipitate; W, whole cell extract at a 1:10 dilution before immunoprecipitation. **e**, *In vitro* kinase activity associated with cyclin D2 (top) and cyclin E1 (bottom) immunoprecipitates using pRb as substrate. M, mock immunoprecipitate.

Cdc2a^{mut1} oocytes and spermatozoa are viable. These observations indicate that Cdk1 is essential for the early stages of embryonic development.

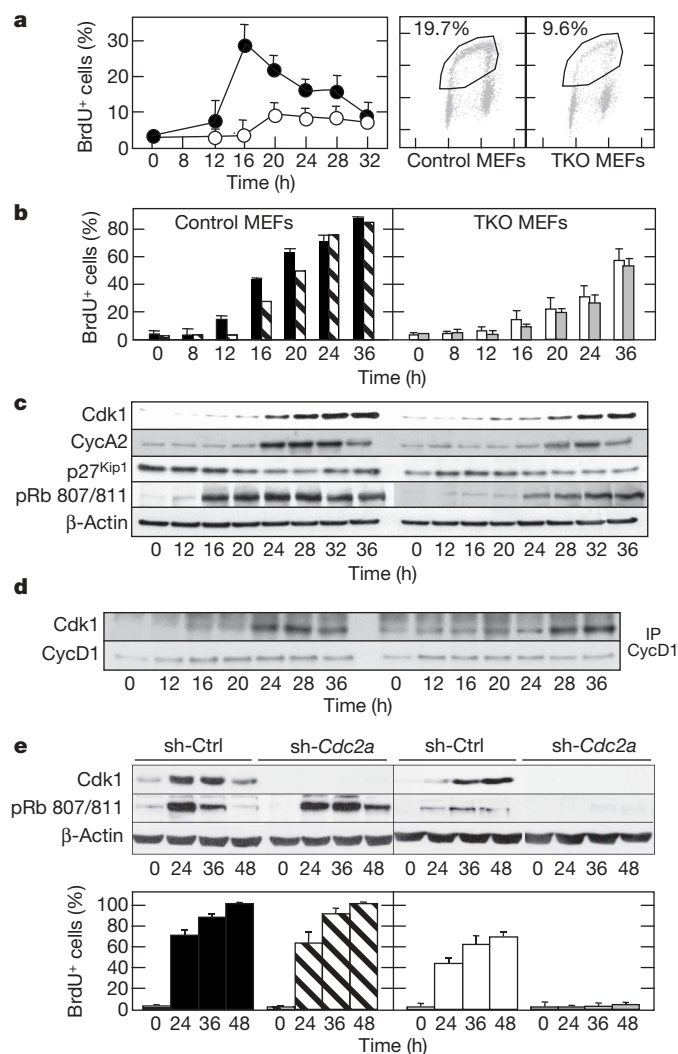


Figure 2 | Mitogenic response of TKO MEFs. **a**, Left, percentage of quiescent *Cdk4*^{+/+}; *Cdk6*^{-/-}; *Cdk2*^{+/+} (filled circles) and *Cdk4*^{-/-}; *Cdk6*^{-/-}; *Cdk2*^{-/-} (open circles) primary MEFs in S phase after serum stimulation following 2-h BrdU pulses. Right, representative analysis of BrdU incorporation 20 h after serum stimulation of the above MEFs. Numbers indicate the percentage of BrdU⁺ cells in S phase. **b**, Percentage of quiescent primary (solid bars) and immortal (hatched bars) *Cdk4*^{+/+}; *Cdk6*^{-/-}; *Cdk2*^{+/+} control MEFs and primary (empty bars) and immortal (grey bars) *Cdk4*^{-/-}; *Cdk6*^{-/-}; *Cdk2*^{-/-} TKO MEFs entering S phase upon addition of 10% FBS in the continuous presence of BrdU at the indicated times after serum stimulation. Data shown as mean \pm s.d., $n = 3$. **c**, Immunoblot analysis of Cdk1, cyclin A2 and p27^{Kip1} expression and phosphorylation of S807/811 pRb residues in immortal MEFs at the indicated times after the addition of serum. Expression of β -actin serves as a loading control. **d**, Formation of Cdk1–cyclin D1 complexes during G0 exit in immortal MEFs shown in **b**. Extracts were immunoprecipitated using antibodies against cyclin D1 and analysed by immunoblotting using antisera against Cdk1. Immunoprecipitation of cyclin D1 is shown as a control. **e**, Knock-down of *Cdk1* abolishes G0 exit in TKO but not in control MEFs. Upper, expression of Cdk1 and phosphorylation of S807/811 pRb residues in *Cdk4*^{+/+}; *Cdk6*^{-/-}; *Cdk2*^{+/+} (left) and *Cdk4*^{-/-}; *Cdk6*^{-/-}; *Cdk2*^{-/-} (right) MEFs infected with lentiviruses expressing control (sh-Ctrl) or *Cdc2a*-specific (sh-*Cdc2a*) shRNAs, at the indicated times after serum stimulation. β -actin serves as a loading control. Lower, percentage of quiescent *Cdk4*^{+/+}; *Cdk6*^{-/-}; *Cdk2*^{+/+} (left) and *Cdk4*^{-/-}; *Cdk6*^{-/-}; *Cdk2*^{-/-} (right) MEFs entering S phase upon addition of 10% FBS in the continuous presence of BrdU infected with lentiviruses expressing a control shRNA (solid and empty bars, respectively) or an shRNA specific for *Cdc2a* (hatched and grey bars, respectively). Data shown as mean \pm s.d., $n = 3$.

Eukaryotic evolution has endowed multicellular organisms with an increasing number of loci encoding Cdks, reaching at least 12 (and potentially 20) in mammals². Unlike in yeasts, where Cdk1 is sufficient to drive the cell cycle, in mammals at least four additional Cdks (Cdk2, Cdk3, Cdk4 and Cdk6) have been proposed to drive the various phases of their cell cycle in sequence^{1,2}. Recent genetic evidence has challenged some of the concepts behind this model. For instance, most laboratory strains of mice are deficient in Cdk3 owing to a naturally occurring mutation³. Moreover, Cdk4 and Cdk6 are not essential for resting cells to exit quiescence⁸. Likewise, cells lacking Cdk2, a kinase that was previously thought to be essential for driving the S phase, divide normally^{6,7}. Finally, MEFs lacking Cdk4 and Cdk2 proliferate and become immortal¹⁰. *In vivo*, single and double Cdk mutant mice show proliferative defects, but only in specific cell types (Fig. 4). Moreover, as illustrated here, embryos

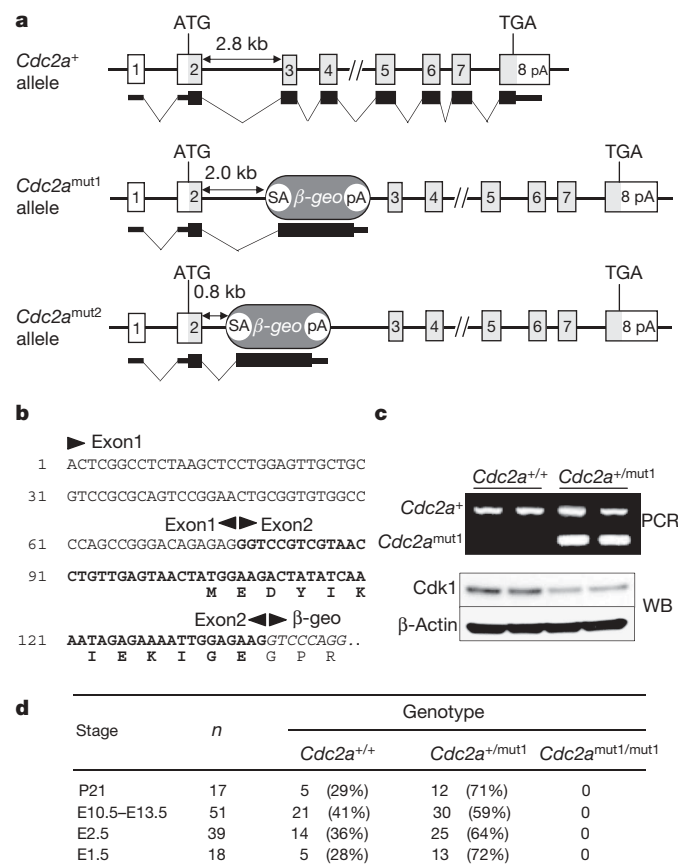


Figure 3 | Cdk1 is essential for early embryonic development. **a**, Schematic representation of wild type *Cdc2a*⁺ (encoding Cdk1), *Cdc2a*^{mut1} and *Cdc2a*^{mut2} alleles. The mutant alleles contain an inserted β -geo trap cassette (dark grey) at different positions within intron 2. *Cdc2a* exons are indicated by numbered boxes. Non-coding (open boxes) and coding (light grey boxes) sequences are indicated. Thin (non-coding) and thick (coding sequences) lines represent the expected mRNAs derived from each allele. ATG, initiation codon; TGA, terminator codon; SA, splicing acceptor; β -geo, bacterial β -galactosidase/neomycin-resistance gene; pA, polyadenylation signal. **b**, Nucleotide sequence of the 5' RACE product derived from *Cdc2a*^{+/mut1} embryonic stem cells. The boundaries of exon 1, exon 2 (bold) and β -geo (italics) are indicated. The amino-acid residues of the predicted Cdk1– β -geo chimaeric protein are also indicated. Bold residues correspond to those of Cdk1. Residues in italics are those encoded by β -geo sequences. An additional glycine residue is expected to be generated during splicing. **c**, Top, genotyping of *Cdc2a*^{+/+} and *Cdc2a*^{+/mut1} mice by PCR amplification of *Cdc2a*⁺ and *Cdc2a*^{mut1} alleles. Bottom, western blot (WB) analysis of Cdk1 protein levels in primary *Cdc2a*^{+/mut1} and wild-type MEFs. **d**, Number of mice (P21) and embryos (E10.5–E13.5, E2.5 and E1.5) of the indicated genotype derived from crosses between *Cdc2a*^{+/mut1} mice. Percentage of *Cdc2a*^{+/+} and *Cdc2a*^{+/mut1} mice and embryos is indicated in parentheses.

lacking all interphase Cdks undergo organogenesis and develop to midgestation, indicating that Cdk1 can drive the essential cell cycle (Fig. 4). Indeed, only mice lacking Cdk1 (this study), cyclin A2 and possibly cyclin B1 have impaired cell division during the very early stages of embryonic development^{21,22}, which indicates that Cdk1 kinase activity cannot be compensated by interphase Cdks.

These observations suggest that Cdks display a significant degree of plasticity. As shown here, Cdk1 can interact with all cyclins, a property that was previously thought to be unique to unicellular organisms. More importantly, these Cdk1 complexes are biochemically active. But these interactions, at least *in vitro*, are only observed in the absence of their 'cognate' Cdk, mainly Cdk4 and Cdk2 (Fig. 1d). Thus, if these preferred Cdk–cyclin interactions also occur *in vivo*, Cdk1 would only drive interphase in the absence of the other Cdks (Fig. 4). A possible resolution to this issue might be provided by strains of mice that express kinase-inactive Cdks rather than null alleles.

Genetic analysis does not explain how Cdk1 can compensate for the absence of interphase Cdks during embryonic development, the most active period of cell division, but not in certain specialized cell types. The

most plausible explanation is that interphase Cdks, but not Cdk1, phosphorylate substrates that are unique to those cell types in which they are essential. It is also possible that in these specialized cells, only interphase Cdks can generate the necessary levels of kinase activity to drive cell division. This scenario might explain why TKO MEFs achieve normal levels of cell proliferation only in the presence of SV40 T₁₂₁, which inactivates the Rb family of proteins (Supplementary Fig. 3c). A precise understanding of how each Cdk contributes to drive cell division should help us to develop appropriate therapeutic strategies to treat cancer as well as other diseases caused by aberrant cell proliferation.

METHODS SUMMARY

Gene targeting and mouse strains. *Cdk4*^{-/-}, *Cdk2*^{-/-}, *Cdk6*^{-/-} and *Cdk2*^{lox} mice have been described^{6,8,10}. *Cdc2a*^{+/-mut1} embryonic stem cells were obtained in a gene trap screen¹⁹ and *Cdc2a*^{+/-mut2} embryonic stem cells were obtained from the German Gene Trap Consortium²⁰.

Cell culture assays. MEFs were isolated from E12.5 embryos. All cell culture experiments were carried out in fibronectin-coated plates (Becton Dickinson) and 3% oxygen. MEFs were propagated according to standard 3T3 protocols. For proliferation assays, we plated 5×10^4 cells on six-well plates in duplicate as described²³. To analyse S phase entry, we deprived MEFs of serum for 72 h in DMEM plus 0.1% FBS and re-stimulated them with 10% FBS. Cells were either pulsed (2 h) or continuously labelled with 50 μ M bromodeoxyuridine, harvested at the indicated times and stained with anti-BrdU fluorescent antibodies. MEFs were infected with retroviral vectors as described²³. Knockdown of *Cdc2a* was achieved with lentiviral shRNA delivery. After infection, cells were selected with 2 μ g ml⁻¹ puromycin for 72 h in DMEM plus 0.1% FBS and re-stimulated with 10% FBS.

Analysis of haematopoietic precursors. Livers were collected from E12.5–E13.5 embryos and disrupted to form single-cell suspensions. Cells were immunostained and haematopoietic progenitors identified as described⁸.

Protein analysis. Protein lysates from either whole embryos or cells were prepared and used for immunoblotting as described²³. For kinase assays, 300 μ g of whole-cell extract was immunoprecipitated and the resulting immunoprecipitates were incubated with 1 μ g of mouse pRb protein fragment or histone H1 as substrates²³. Cdk–cyclin complexes were analysed by immunoprecipitation of 300 μ g of whole-embryo extracts.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 May; accepted 21 June 2007.

- Bloom, J. & Cross, F. R. Multiple levels of cyclin specificity in cell-cycle control. *Nature Rev. Mol. Cell Biol.* **2**, 149–160 (2007).
- Malumbres, M. & Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* **30**, 630–641 (2005).
- Ye, X., Zhu, C. & Harper, J. W. A premature-termination mutation in the *Mus musculus* cyclin-dependent kinase 3 gene. *Proc. Natl Acad. Sci. USA* **98**, 1682–1686 (2001).
- Rane, S. G. *et al.* Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in β -cell hyperplasia. *Nature Genet.* **22**, 44–52 (1999).
- Tsutsui, T. *et al.* Targeted disruption of Cdk4 delays cell cycle entry with enhanced p27^{Kip1} activity. *Mol. Cell Biol.* **19**, 7011–7019 (1999).
- Ortega, S. *et al.* Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice. *Nature Genet.* **35**, 25–31 (2003).
- Berthet, C., Aleem, E., Coppola, V., Tassarollo, L. & Kaldis, P. Cdk2 knockout mice are viable. *Curr. Biol.* **13**, 1775–1785 (2003).
- Malumbres, M. *et al.* Mammalian cells cycle without the D-type Cyclin-dependent kinases Cdk4 and Cdk6. *Cell* **118**, 493–504 (2004).
- Berthet, C. *et al.* Combined loss of Cdk2 and Cdk4 results in embryonic lethality and Rb hypophosphorylation. *Dev. Cell* **10**, 563–573 (2006).
- Barriere, C. *et al.* Mice thrive without Cdk4 and Cdk2. *Mol. Oncology* **1**, 72–83 (2007).
- Kozar, K. *et al.* Mouse development and cell proliferation in the absence of D-cyclins. *Cell* **118**, 477–491 (2004).
- Geng, Y. *et al.* Kinase-independent function of cyclin E. *Mol. Cell* **25**, 127–139 (2007).
- Geng, Y. *et al.* Cyclin E ablation in the mouse. *Cell* **114**, 431–443 (2003).
- Parisi, T. *et al.* Cyclins E1 and E2 are required for endoreplication in placental trophoblast giant cells. *EMBO J.* **22**, 4794–4803 (2003).
- Zarkowska, T. & Mitnacht, S. Differential phosphorylation of the retinoblastoma protein by G1/S cyclin-dependent kinases. *J. Biol. Chem.* **272**, 12738–12746 (1997).
- Aleem, E., Kiyokawa, H. & Kaldis, P. Cdc2-Cyclin E complexes regulate the G1/S phase transition. *Nature Cell Biol.* **7**, 831–836 (2005).

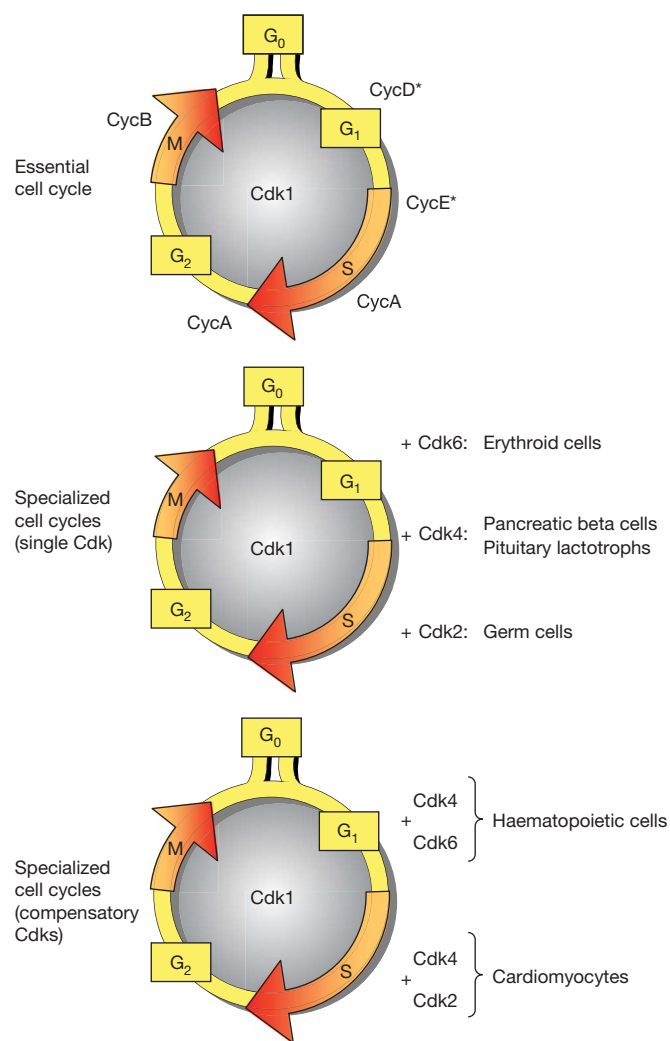


Figure 4 | Summary of genetic results obtained by ablation of loci encoding cell-cycle Cdks in mice. The diagram depicting the 'essential cell cycle' (top) is based on our results. The asterisks indicate that D-type or E-type cyclins are not essential for cell proliferation^{11–14}. It is not known whether cells expressing Cdk1 in the absence of D-type and/or E-type cyclins would proliferate. The cartoons depicting 'specialized cell cycles' summarize previous work illustrating which interphase Cdk (middle) or Cdks (bottom) are essential for proliferation of the indicated cell types. The location of Cdks within the phases of the cell cycle is arbitrary. Cyclins have been omitted for clarity. References can be found in the text.

17. Saenz-Robles, M. T., Symonds, H., Chen, J. & Van Dyke, T. Induction versus progression of brain tumor development: differential functions for the pRB- and p53-targeting domains of simian virus 40 T antigen. *Mol. Cell. Biol.* **14**, 2686–2698 (1994).
18. Fung, T. K. & Poon, R. Y. A roller coaster ride with the mitotic cyclins. *Semin. Cell Dev. Biol.* **16**, 335–342 (2005).
19. Sutherland, H. G. *et al.* Large-scale identification of mammalian proteins localized to nuclear sub-compartments. *Hum. Mol. Genet.* **10**, 1995–2011 (2001).
20. Schnutgen, F. *et al.* Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc. Natl Acad. Sci. USA* **102**, 7221–7226 (2005).
21. Murphy, M. *et al.* Delayed early embryonic lethality following disruption of the murine cyclin A2 gene. *Nature Genet.* **15**, 83–86 (1997).
22. Brandeis, M. *et al.* Cyclin B2-null mice develop normally and are fertile whereas cyclin B1-null mice die in utero. *Proc. Natl Acad. Sci. USA* **95**, 4344–4349 (1998).
23. Martin, A. *et al.* Cell cycle inhibition and tumor suppression by p21^{Cip1} and p27^{Kip1} are independent of Cdk2. *Cancer Cell* **7**, 591–598 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is dedicated to the memory of our colleagues L. Gómez García and J. López Navarro. We thank R. González, M. San Román, B. Velasco and R. Villar for technical assistance, A. García for help with the

cytometer, and A. Swat and S. Mouron for advice on quantitative RT-PCR. We also value the support provided by the Transgenic and Comparative Pathology Units of the CNIO. We thank M. Puyol for the Cdk2 and Cdk6 antibodies. This work was supported by grants from the Plan Nacional de Investigación Científica (D.S., M.M. and M.B.), the OncoCycle programme from the Comunidad de Madrid (M.M. and M.B.), Fondo de Investigación Sanitaria (D.S.), V Framework Programme of the European Union (M.B.) and INSERM and Association pour la Recherche contre le Cancer (P.D.). C.T., C.B. and A.C. were supported by fellowships from FEBS, la Ligue contre le Cancer (Comité de la Dordogne) and FPI (Ministerio de Educación y Ciencia), respectively.

Author Contributions M.B. supervised the entire project. M.B., D.S. and M.M. conceived and designed the experiments, and wrote the manuscript with comments from co-authors. C.B. generated the mouse strains and carried out most cell culture experiments and part of the protein analysis. A.C. was responsible for most protein analysis and part of the cell culture experiments. S.H. and C.T. characterized *Cdc2a* mutant mice. P.D. was responsible for histopathological analysis. J.F.C. and K.N. generated one of the *Cdc2a* mutant embryonic stem cell clones.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.B. (mbarbacid@cnio.es).

METHODS

Gene targeting and mouse strains. *Cdk4*^{-/-}, *Cdk2*^{-/-}, *Cdk6*^{-/-} and *Cdk2*^{lox} mice have been described^{6,8–10}. *Cdc2a*^{+/-mut1} embryonic stem (ES) cells were obtained in a gene trap screen¹⁹. Neo^R colonies were screened for β-galactosidase activity with 5-bromo-4-chloro-3-indolyl β-D-galactosidase (X-Gal) and analysed by immunofluorescence using antibodies specific for β-gal. To characterize the trapped allele, we prepared total RNA using Bio/RNA-X-cell (Bio/Gene Limited). 5' RACE products were generated and directly sequenced using Big DyeTM terminator kit using a -40 lacZ primer (USB/Amersham) on an ABI prism 377 DNA sequencer. *Cdc2a*^{+/-mut2} ES cells were obtained from the German Gene Trap Consortium (clone P009F08; <http://www.genetrap.org>). The sequence of the oligonucleotides used for genotyping is available upon request. Mice used in this study have been maintained in a mixed 129/SvJ x C57BL/6J background according to the Guide of the Care and Use of Laboratory Animals (ILAR 1996) and in accordance with the Spanish Animal Protection Law (RD1201/2005) and the European Directive (86/609/CEE) established by the European Union to regulate animal care standards.

Histopathology and immunohistochemistry. Embryos were fixed in 10%-buffered formalin (Sigma) and embedded in paraffin. Three- or five-micrometer-thick sections were stained with haematoxylin and eosin. For proliferation studies, tissue sections were stained with Ki67-specific antibodies (MIB-1; Dako). Apoptotic cells were detected using anti-active Caspase 3 antibodies (R&D Systems).

Cell culture assays. MEFs were isolated from E12.5 embryos and cultured in fibronectin-coated plates (Becton Dickinson) and 3% oxygen in the presence of Dulbecco's modified Eagle's medium (DMEM) supplemented with 2 mM glutamine, 1% penicillin/streptomycin and 10% fetal bovine serum (FBS). MEFs were propagated according to standard 3T3 protocols. For proliferation assays, 5×10^4 cells were plated on six-well plates in duplicate as described²³. To analyse S phase entry, we deprived MEFs (10^6 cells per 10-cm dish) of serum for 72 h in DMEM plus 0.1% FBS and re-stimulated with 10% FBS. Cells were either pulsed (2 h) or continuously labelled with 50 μM bromodeoxyuridine (BrdU; Sigma), harvested at the indicated times and stained with anti-BrdU fluorescent antibodies (Becton Dickinson). MEFs were infected with retroviral vectors as described²³. Knockdown of *Cdc2a* was mediated with lentiviral Mission-shRNA plasmids (SHGLY-NM_007659, Sigma) according to the manufacturer's instructions. After infection, cells were selected with $2 \mu\text{g ml}^{-1}$ puromycin for 72 h in DMEM plus 0.1% FBS and re-stimulated with 10% FBS.

Analysis of haematopoietic precursors. Livers were collected from E12.5–E13.5 embryos and disrupted to form single-cell suspensions. Cells were immunostained and haematopoietic progenitors identified as described⁸. Haematopoietic stem cells were gated as Lin⁻IL7Rα⁻c-Kit⁺Sca1⁺; common myeloid progenitors as Lin⁻IL7Rα⁻c-Kit⁺Sca1⁻FcγR^{low}CD34⁺; granulocyte-macrophage progenitors as Lin⁻IL7Rα⁻c-Kit⁺Sca1⁻FcγR^{hi}CD34⁺; and megakaryocyte-erythroid progenitors as Lin⁻IL7Rα⁻c-Kit⁺Sca1⁻FcγR^{low}CD34⁻. Their relative numbers were quantified using a FACSaria (Becton Dickinson) cytometer.

Protein analysis. Protein lysates were prepared and used for immunoblotting as described²³. Antibodies against the following proteins were used: Cdk1 (17; Santa Cruz), Cdk4 (C22; Santa Cruz), Cdk2 and Cdk6 (our own rabbit polyclonal antibodies), Cdk7 (C22; Santa Cruz), Cdk9 (C22; Santa Cruz), Cyclin A2 (H432; Santa Cruz), Cyclin B1 (Santa Cruz), Cyclin D1 (DCS6 and AB4; Neo Markers), Cyclin D2 (Santa Cruz), Cyclin E1 (M20; Santa Cruz), p21^{Cip1} (C19; Santa Cruz), p27^{Kip1} (Transduction Laboratories), p57^{Kip2} (C20; Santa Cruz), β-Actin (AC15, Sigma) and pRb (BD Pharmingen). pRb phosphospecific antibodies to phosphorylated residues S608 (#2181), S780 (#9307) and S807/811 (#9308) were from Cell Signaling. As secondary antibodies, we used peroxidase-conjugated IgG (Dako), followed by chemiluminescence detection (ECL; Amersham). For kinase assays, 300 μg of whole cell extract was immunoprecipitated with the indicated antibodies and the resulting immunoprecipitates incubated with 1 μg of mouse pRb protein fragment (amino-acid residues 769–921; Santa Cruz) or histone H1 (calf thymus, Roche) as substrates²³.

***Cdkn1b* expression analysis by real-time quantitative PCR.** Total RNA was isolated from cells using RNAeasy (Quiagen) and digested with DNaseI to remove contaminating DNA. cDNA was prepared from 2 μg of total RNA using Superscript II Reverse transcriptase kit (Invitrogen). Real-time PCR was performed on several cDNA dilutions plus 1x SYBR Green PCR Master Mix (Applied Biosystems) and 50 nM of *Cdkn1b* primers (sequence available upon request). Reactions were carried out using an ABI Prism 7500 Sequence Detection System (Applied Biosystems) for 40 cycles (95 °C for 15 s and 60 °C for 1 min) after an initial 10-min incubation at 95 °C. Variation of *Cdkn1b* expression was calculated using the ΔΔCt method²⁴ with *β-actin* mRNA as an internal control.

24. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ method. *Methods* **25**, 402–408 (2001).

LETTERS

Small self-RNA generated by RNase L amplifies antiviral innate immunity

Krishnamurthy Malathi¹, Beihua Dong¹, Michael Gale Jr² & Robert H. Silverman¹

Antiviral innate immunity is initiated in response to RNA molecules that are produced in virus-infected cells¹. These RNAs activate signalling cascades that activate the genes that encode α - and β -interferon (IFN). Signalling occurs through the interaction of the RNAs with either of two pathogen recognition receptors, retinoic acid-inducible gene-I (RIG-I, also known as DDX58) and melanoma differentiation associated gene-5 (MDA5, also known as IFIH1), which contain amino-terminal caspase activation and recruitment domains (CARD) and carboxy-terminal DExD/H Box RNA helicase motifs^{2–5}. RIG-I and MDA5 interact with another CARD protein, interferon- β promoter stimulator protein-1 (IPS-1, also known as MAVS, VISA and Cardif), in the mitochondrial membrane, which relays the signal through the transcription factors interferon regulatory factor 3 (IRF-3) and nuclear factor (NF)- κ B to the IFN- β gene^{6–10}. Although the signalling pathway is well understood, the origin of the RNA molecules that initiate these processes is not. Here we show that activation of the antiviral endoribonuclease, RNase L¹¹, by 2',5'-linked oligoadenylate (2-5A)¹² produces small RNA cleavage products from self-RNA that initiate IFN production. Accordingly, mouse embryonic fibroblasts lacking RNase L were resistant to the induction of IFN- β expression in response to 2-5A, dsRNA or viral infection. Single-stranded regions of RNA are cleaved 3' of UpUp and UpAp sequences by RNase L during viral infections, resulting in small, often duplex, RNAs^{13,14}. We show that small self-RNAs produced by the action of RNase L on cellular RNA induce IFN- β expression and that the signalling involves RIG-I, MDA5 and IPS-1. Mice lacking RNase L produce significantly less IFN- β during viral infections than infected wild-type mice. Furthermore, activation of RNase L with 2-5A *in vivo* induced the expression of IFN- β in wild-type but not RNase L-deficient mice. Our results indicate that RNase L has an essential role in the innate antiviral immune response that relieves the requirement for direct sensing of non-self RNA.

To determine the effect of RNase L on the induction of IFN- β , we transfected wild-type and *RnaseL*^{-/-} (RNase L-deficient)¹⁵ mouse embryonic fibroblasts (MEFs) with either dephosphorylated (inactive) or triphosphorylated (active) trimer 2-5A¹⁶. Transfection of dephosphorylated 2-5A ((2'-5')A₃) did not induce the expression of IFN- β , as determined by a specific enzyme-linked immunosorbent assay (ELISA; Fig. 1a). However, triphosphorylated 2-5A ((2'-5')p₃A₃), the naturally occurring activator of RNase L, potently induced IFN- β production in a dose-dependent manner in wild-type but not in RNase L-deficient cells. The RNase L-deficient cells were also relatively resistant to the induction of IFN- β expression in response to the synthetic dsRNA poly(I):poly(C) in comparison to the wild-type cells (Fig. 1b). In addition, RNase L-deficient cells were partially resistant to the induction of IFN- β expression by the paramyxovirus, Sendai virus

(SeV), a negative RNA strand virus (infection of wild-type cells produced 7.5-fold more IFN- β than did infection of RNase L-deficient cells; Fig. 1c). Ectopic expression of RNase L in RNase L-deficient MEFs restored the induction of IFN- β expression by SeV, poly(I):poly(C) or 2-5A (Supplementary Fig. 1). Showing that the nuclease function of RNase L is essential for signalling, the nuclease-dead mutants of RNase L (R667A or H672A) failed to restore the induction of IFN- β expression by 2-5A (Supplementary Fig. 1c, d).

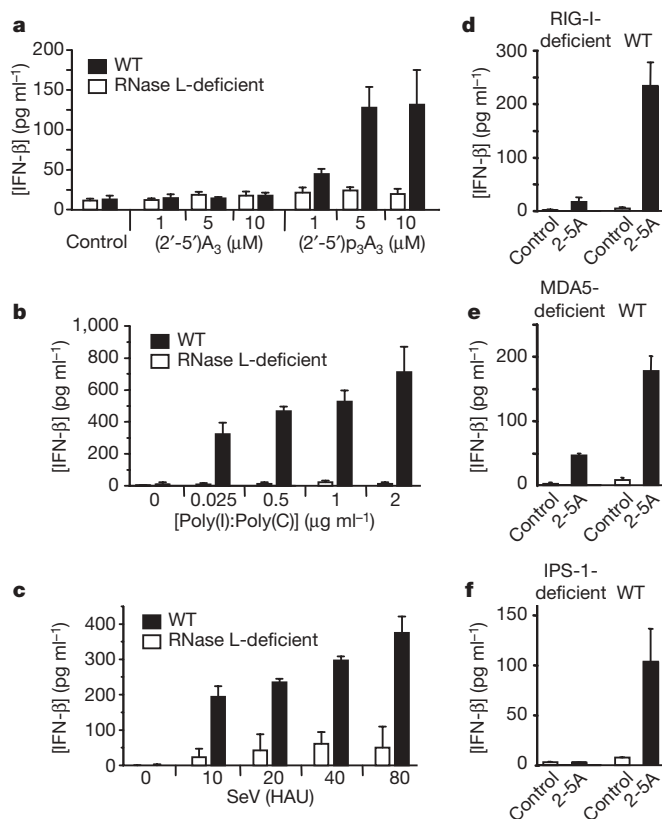


Figure 1 | Involvement of RNase L in induction of IFN- β expression by 2-5A, dsRNA or viral infection. **a–c**, IFN- β concentration from wild-type (WT) or RNase L-deficient MEFs that were mock transfected (Control) or transfected with (2'-5')A₃ or (2'-5')p₃A₃ for 16 h (**a**); mock transfected or transfected with poly(I):poly(C) for 18 h (**b**); or mock infected or infected with SeV for 18 h (**c**). **d–f**, IFN- β concentration from wild-type MEFs and RIG-I-deficient MEFs (**d**), MDA5-deficient MEFs (**e**) or IPS-1-deficient MEFs that were either mock transfected or transfected with unfractionated 2-5A for 16 h (**f**). Values are means from triplicate assays with standard deviations (s.d.).

¹Department of Cancer Biology, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. ²Department of Immunology, School of Medicine University of Washington, 1959 NE Pacific Street, H-578 Health Sciences, Box 357650 Seattle, Washington 98195-7650, USA.

To test our hypothesis that RNA cleavage products generated by RNase L were responsible for the induction of IFN- β expression, we performed experiments with primary MEFs lacking the RNA helicases RIG-I or MDA5, or their adaptor, IPS-1. The induction of IFN- β in response to treatment with 2-5A was greatly reduced (by >14-fold) in *Rig-i*^{-/-} (RIG-I-deficient)³ MEFs as compared to wild-type MEFs (Fig. 1d). In addition, *Mda5*^{-/-} (MDA5-deficient) MEFs⁴ were about fourfold less responsive to the induction of IFN- β by 2-5A than were wild-type MEFs (Fig. 1e). Both RIG-I and MDA5 relay signals through IPS-1⁶⁻⁹. As a result, *Ips1*^{-/-} (IPS-1-deficient) MEFs¹⁷ were unresponsive to the induction of IFN- β by 2-5A (Fig. 1f).

To extend these findings to human cells, we used small inhibitory (si)RNA oligonucleotides to suppress individually the levels of RIG-I, MDA5 and IPS-1 in the human prostate cancer cell line DU145 (Fig. 2a). A decrease in the level of each of these signalling proteins inhibited the induction of a human IFN- β promoter by 2-5A (by more than sevenfold compared to the untreated control; Fig. 2b). By contrast, non-specific short interfering (si)RNA failed to inhibit the induction of the IFN- β promoter by 2-5A. Simultaneous knock-down of RIG-I and MDA5 nearly prevented the induction of the IFN- β promoter by 2-5A. By contrast, a reduction in the level of the RNA helicase laboratory of genetics and physiology 2 (LGP2)¹⁸ slightly increased signalling to the IFN- β promoter by 2-5A (Supplementary Fig. 2). In addition, 2-5A-induction of the IFN- β promoter in the human hepatoma cell line Huh7.5, which has a mutation in RIG-I (T551)¹⁹, was 28% that of 2-5A-induced Huh7 cells, which contain wild-type RIG-I (Fig. 2c). The impact of the RIG-I mutation on autocrine IFN signalling was apparent from the lower levels of phosphorylation of signal transducer and activator of transcription-1 (STAT-1) after 2-5A treatment in the Huh7.5 cells compared with identically treated Huh7 cells (Fig. 2d, compare lanes 3 and 6).

To provide direct evidence that RNA cleavage products generated by RNase L are responsible for signalling, we isolated total cellular

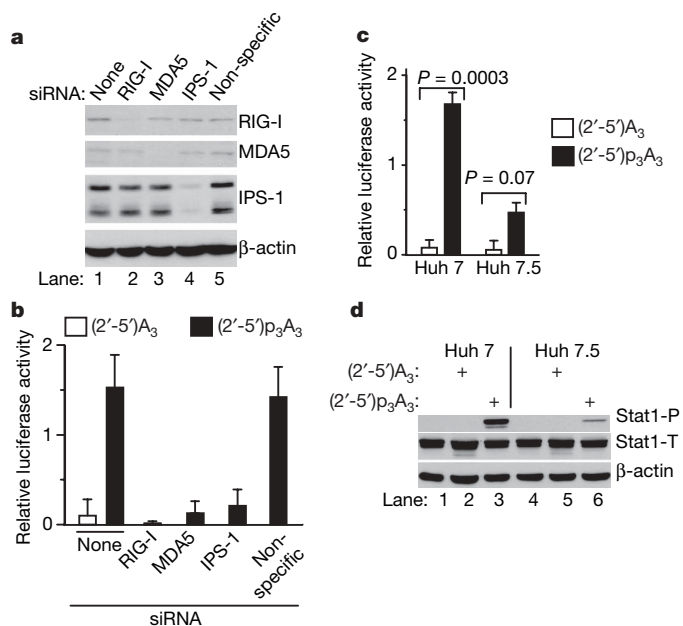


Figure 2 | 2-5A induces transcriptional activation of the IFN- β promoter. **a**, RIG-I, MDA5, IPS-1 and β -actin in DU145 cells treated for 48 h with siRNAs (shown above) in immunoblots. **b**, Activation of the IFN- β promoter in DU145 cells treated with siRNAs (5 nM) and 1.0 μ M of (2'-5')A₃ or (2'-5')p₃A₃. **c**, Activation of the IFN- β promoter in Huh7 or Huh7.5 cells transfected with 1.0 μ M of (2'-5')A₃ or (2'-5')p₃A₃ for 18 h. Relative luciferase (LUC) activity, firefly LUC/*Renilla* LUC. **d**, Stat1 phosphorylation in Huh7 or Huh7.5 cells treated with (2'-5')p₃A₃ or (2'-5')A₃ for 18 h. Stat1-P, phosphorylated Stat1; Stat1-T, total Stat1. Error bars, s.d.

RNA from RNase L-deficient MEFs and incubated it with recombinant, purified RNase L in the presence and absence of 2-5A. The reactions were allowed to proceed until no further RNA cleavage could be detected, as monitored by fluorescence resonance energy transfer (FRET) assays using an internal RNA FRET probe²⁰ and by RNA size analysis using RNA chips (Fig. 3a and Supplementary Fig. 3). The production of IFN- β by MEFs increased (to sixfold higher levels) after they were transfected with RNA pre-incubated with RNase L plus 2-5A, as compared to the production of IFN- β after transfection with RNA incubated in the absence of RNase L and 2-5A (Fig. 3b). Incubation of RNA with RNase L without 2-5A provided a much smaller (twofold) increase in IFN- β induction when the RNA was subsequently transfected into MEFs, probably owing to a low level of RNA cleavage under these conditions (Supplementary Fig. 3). The RNA cleavage products that were <200 nucleotides long were isolated by a solid-phase fractionation method (Fig. 3a, lanes 5–7; see Methods). Using equivalent amounts of total RNA as substrate, IFN- β was induced to >10-fold higher levels in response to transfection of MEFs with small RNAs produced by incubation with RNase L plus 2-5A compared to transfection with small RNAs isolated from an

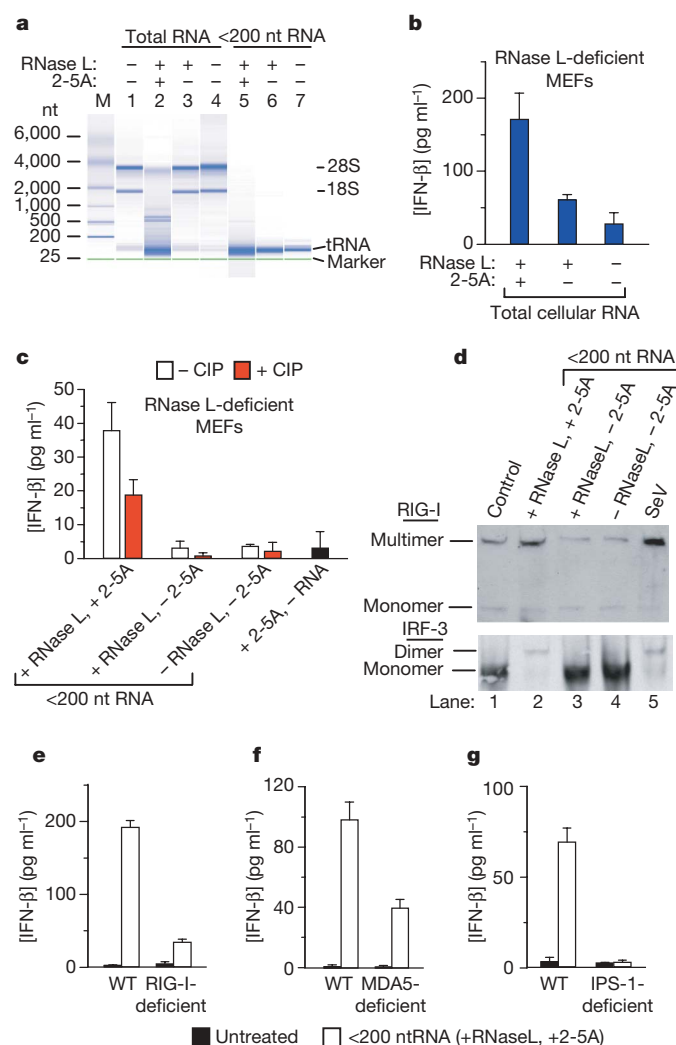


Figure 3 | Cleavage of cellular RNA by RNase L produces small RNAs that activate the IFN- β gene. **a**, RNA separated in RNA chip. **b**, **c**, IFN- β from MEFs transfected (for 18 h) with total (10 μ g) RNA (**b**) or <200-nucleotide (nt) RNA (from 20 μ g total RNA) (**c**). **d**, RIG-I (upper) and IRF-3 (lower) in RNase L-deficient MEFs and HT1080 cells, respectively, with <200-nt RNA (lanes 1–4) or wild-type MEFs (upper) or HT1080 cells (lower) infected with SeV (for 18 h) (lane 5). **e**–**g**, Levels of IFN- β from MEFs (wild-type, RIG-I-deficient, MDA5-deficient or IPS-1-deficient) incubated for 18 h with or without <200-nt RNA. Error bars, s.d.

equivalent amount of total RNA incubated with or without RNase L alone (Fig. 3c). Small RNA at $\leq 50 \text{ ng ml}^{-1}$ produced by incubation with 2-5A-activated RNase L induced IFN- β and its promoter in MEFs and in human HT1080 cells, respectively (Supplementary Fig. 4). By contrast, up to $1 \mu\text{g ml}^{-1}$ of the small RNA fraction obtained in the absence of RNase L digestion had little or no effect.

It has been reported that the presence of free 5' triphosphorylated or duplex structures in RNA serve to discriminate self (cellular) from non-self (viral) RNA substrates and is used by RIG-I to recognise viral RNA and to initiate innate immune signalling^{21,22}. However, the small RNAs produced by RNase L contain 3', and not 5', monophosphoryl groups¹³. To determine how the 3'-phosphoryls might contribute to the induction of IFN expression, we treated the small RNA cleavage products with calf intestinal phosphatase (CIP) under conditions that completely dephosphorylated an internal control radiolabelled RNA substrate (data not shown). Dephosphorylation with CIP reduced the production of IFN- β by 50% (Fig. 3c). These results indicate that RIG-I and/or MDA5 are stimulated by 3'-phosphorylated RNAs. However, the small RNAs that lacked terminal phosphates were still capable of signalling, albeit at a reduced level, perhaps owing to duplex structures.

RIG-I forms multimeric complexes and IRF-3 dimerizes during signalling initiated by SeV infections, as determined by native polyacrylamide gel electrophoresis (PAGE) and immunoblotting²³ (see also Fig. 3d, lane 5). Similarly, RIG-I formed multimers and IRF-3 dimerized in response to the small RNAs (produced by RNase L plus 2-5A) in RNase L-deficient MEFs and HT1080 cells, respectively (Fig. 3d, lanes 2). However, RNAs less than 200 nucleotides long that were isolated after incubation of total RNA with RNase L in the absence of 2-5A or without RNase L failed to induce either RIG-I multimers or IRF-3 dimers (Fig. 3d, lanes 3 and 4). Furthermore, IFN- β was potently induced in wild-type MEFs in response to treatment with the small RNAs (Fig. 3e–g). By contrast, IFN- β production by RIG-I-deficient and MDA5-deficient cells transfected with the small RNAs was 17.5% and 40% of the levels obtained in identically treated wild-type cells, whereas there was no detectable IFN- β produced in IPS-1-deficient cells that had been treated with the small RNAs. These results indicate that the small RNAs signal through RIG-I and MDA5 to IPS-1, thereby activating IRF-3.

To establish how RNase L contributes to IFN- β production *in vivo*, we infected wild-type and RNase L-deficient mice with the picornavirus, encephalomyocarditis virus (EMCV), a positive RNA strand virus or SeV. Infection of wild-type mice with EMCV produced, on average, 2.2- and 3.5-fold greater levels of IFN- β when measured 6 and 24 h after infection, respectively, when compared with similarly infected RNase L-deficient mice (Fig. 4a). The effect of RNase L was even more pronounced after SeV infection: wild-type mice produced, on average, sixfold greater levels of IFN- β when measured 6 h after infection than did RNase L-deficient mice (Fig. 4b). The increased IFN production in the wild-type mice, compared with the RNase L-deficient mice, was highly significant with both types of virus.

To establish that self-RNA cleavage products could signal to the IFN- β gene *in vivo*, we injected 2-5A into uninfected mice by the intraperitoneal route. Mock-treated wild-type and RNase L-deficient mice showed no induction of IFN- β . There was also no induction of IFN- β by 2-5A in the RNase L-deficient mice. In contrast, 2-5A induced circulating IFN- β in all wild-type mice ($n = 6$; Fig. 4c).

These findings show that RNase L is crucial for enhancing IFN- β production through the RIG-I–MDA5–IPS-1 cascade. Previously, we showed that activation of RNase L by 2-5A induced the expression of twice as many messenger RNA species, including mRNAs for many IFN- β -stimulated genes, as it suppressed²⁴. The current study provides evidence that transcriptional signalling is initiated by RNase L-generated RNA cleavage products of cellular (self)-RNA (Fig. 4d). Viruses produce 'non-self' RNAs consisting of double-stranded (ds)RNA and 5'-phosphorylated RNAs that signal through RIG-I

and/or MDA5 to IFN- β genes^{21–23}. In addition, viral dsRNA directly activates 2'-5'-oligoadenylate synthetase (OAS), resulting in the production of 2-5A from ATP¹². 2-5A stimulates RNase L and leads to the production of small RNA cleavage products that terminate in 3'-phosphoryls from cellular self-RNA, as well as from some viral RNAs¹⁴. The products of RNase L cleavage of cellular and viral RNAs probably perform similar functions. Because RNase L cleaves exclusively single-stranded regions in RNA, the cleavage products are often duplex structures. These RNAs activate signalling by RIG-I and MDA5 to IRF-3 through the IPS-1 adaptor protein and its signalling partners, which results in the production of IFN. OAS, RIG-I and MDA5 are all induced by IFN- β at a transcriptional level, thereby further amplifying the production of IFN- β (Fig. 4d; IFN- β -induced proteins are shown in red). Negative regulators that prevent the continuous auto-amplification of IFN- β include 2'-phosphodiesterase, which degrades 2-5A²⁵, and the IFN- β -inducible inhibitors LGP2 (ref. 18) and ubiquitin ligase RNF125 (ref. 26). Although the idea is not explored here, it is also possible that small RNAs generated from RNase L activity could activate toll-like receptor (TLR)3-dependent signalling to drive further IFN- β production¹.

The effect of RNase L on viral infections *in vivo* has been established in studies with RNase L-deficient mice. The ability of mice to survive various viral infections, including EMCV, West Nile virus and Coxsackievirus B4, is compromised by the absence of RNase L^{15,27,28}. In addition, an antiviral role for RNase L in humans is evident from the increased rate of prostate infections by the retrovirus, XMRV, in men that are homozygous for a reduced activity variant of RNase L²⁹. RNase L contributes to the IFN antiviral response by directly cleaving viral and cellular RNA. However, it is now apparent that by generating small RNAs that induce IFN, the effects of the OAS–RNase L pathway extend beyond initially infected cells to

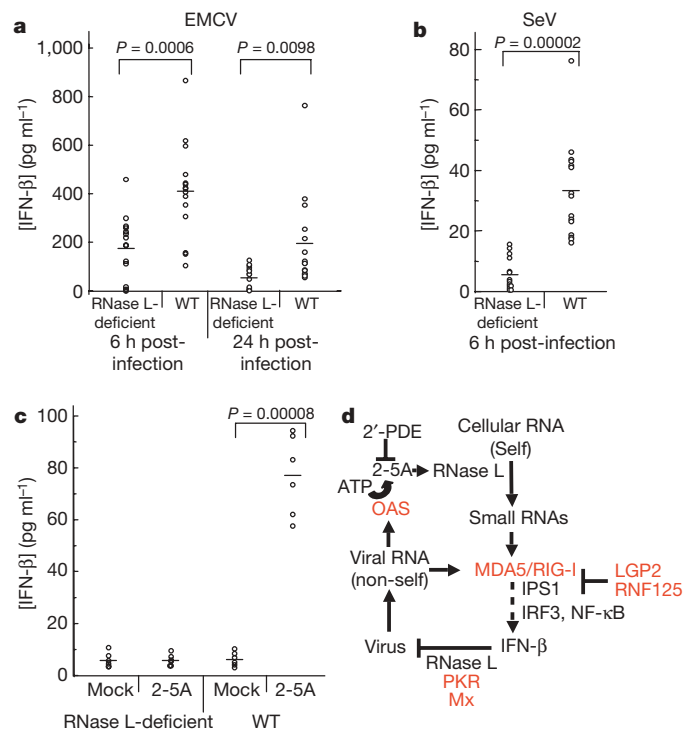


Figure 4 | RNase L contributes to induction of IFN- β by virus or 2-5A *in vivo*. Serum levels of IFN- β from RNase L-deficient or wild-type mice infected with EMCV (**a**) or SeV for the indicated times (**b**), or injected intraperitoneally with 2-5A or mock-treated for 5 h (**c**). Horizontal lines show mean levels of IFN- β . Two-tailed, paired Student's *t*-tests were performed. **d**, Model for transcriptional activation of the IFN- β gene by small RNAs produced by cleavage of cellular (self) RNA with RNase L. Red, IFN-inducible proteins; 2'-PDE, 2'-phosphodiesterase; OAS, 2'-5'-oligoadenylate synthetase.

support a broader antiviral state in the organism. Furthermore, RNase L circumvents the need for non-self viral RNAs in perpetuating and amplifying innate immunity, as by cleaving cellular RNA it produces RNA activators of RIG-I and MDA5. Because many pathogenic viruses have mechanisms to block RIG-I signalling^{9,10} or OAS function³⁰ during infection, strategies to regulate RNase L activation could have direct immunoenhancing and antiviral therapeutic applications.

METHODS SUMMARY

We cleaved total RNA from RNase L-deficient MEFs with RNase L activated by unfractionated 2-5A in the presence of a FRET RNA probe²⁰. Control reactions lacked RNase L and/or 2-5A. Small (<200 nucleotide) RNAs were isolated using the mirVana miRNA Isolation Kit (Ambion). Small RNAs were incubated without or with CIP (NEB). Transfections of 2-5A, RNA or plasmids were performed with lipofectamine 2000 (Invitrogen) whereas poly(I):poly(C) was transfected with Fugene 6 (Roche Applied Science). We transfected siRNA oligonucleotides (5 nM) targeting human RIG-I, MDA5, IPS1, LGP2 or non-specific oligonucleotides (Dharmacon) into DU145 cells with DharmaFECT 1. Cells were transfected after 48 h with luciferase-reporter plasmids and after another 24 h with 2-5A. Luciferase activity was measured 18 h later. Alternately, 24 h after transfecting with the luciferase plasmids, cells were transfected with 2-5A or RNA for 18 h. SeV in serum-free media was added to cells in six-well plates. After 1 h, the medium was replaced with growth medium for an additional 18 h. Wild-type and *Rnasel*^{-/-} mice (*n* = 15 each, 5–6 weeks old) on a C57BL/6 background were inoculated intraperitoneally with 1×10^3 plaque forming units (pfu) of EMCV or intranasally with 320 hemagglutination units (HAU) of SeV. Sera were collected (at the times indicated) after injection and IFN- β levels were determined by ELISA (PBL Biomedical Labs). Control mice received only PBS. Treatment with 2-5A involved intraperitoneal injections of 2-5A and fucose 6 for 5 h. Mock treatments were done with fucose 6 alone.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 May; accepted 21 June 2007.

Published online 25 July 2007.

- Saito, T. & Gale, M. Jr Principles of intracellular viral recognition. *Curr. Opin. Immunol.* **19**, 17–23 (2007).
- Yoneyama, M. et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
- Kato, H. et al. Cell type-specific involvement of RIG-I in antiviral response. *Immunity* **23**, 19–28 (2005).
- Gitlin, L. et al. Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc. Natl Acad. Sci. USA* **103**, 8459–8464 (2006).
- Kato, H. et al. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441**, 101–105 (2006).
- Kawai, T. et al. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* **6**, 981–988 (2005).
- Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF- κ B and IRF 3. *Cell* **122**, 669–682 (2005).
- Xu, L. G. et al. VISA is an adaptor protein required for virus-triggered IFN- β signaling. *Mol. Cell* **19**, 727–740 (2005).
- Meylan, E. et al. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
- Loo, Y. M. et al. Viral and therapeutic control of IFN- β promoter stimulator 1 during hepatitis C virus infection. *Proc. Natl Acad. Sci. USA* **103**, 6001–6006 (2006).
- Zhou, A., Hassel, B. A. & Silverman, R. H. Expression cloning of 2-5A-dependent RNAase: a uniquely regulated mediator of interferon action. *Cell* **72**, 753–765 (1993).
- Kerr, I. M. & Brown, R. E. pppA2'p5'A2'p5'A: an inhibitor of protein synthesis synthesized with an enzyme fraction from interferon-treated cells. *Proc. Natl Acad. Sci. USA* **75**, 256–260 (1978).
- Wreschner, D. H., McCauley, J. W., Skehel, J. J. & Kerr, I. M. Interferon action—sequence specificity of the ppp(A2'p)nA-dependent ribonuclease. *Nature* **289**, 414–417 (1981).
- Han, J. Q., Wroblewski, G., Xu, Z., Silverman, R. H. & Barton, D. J. Sensitivity of hepatitis C virus RNA to the antiviral enzyme ribonuclease L is determined by a subset of efficient cleavage sites. *J. Interferon Cytokine Res.* **24**, 664–676 (2004).
- Zhou, A. et al. Interferon action and apoptosis are defective in mice devoid of 2',5'-oligoadenylate-dependent RNase L. *EMBO J.* **16**, 6355–6363 (1997).
- Dong, B. et al. Intrinsic molecular activities of the interferon-induced 2-5A-dependent RNase. *J. Biol. Chem.* **269**, 14153–14158 (1994).
- Kumar, H. et al. Essential role of IPS-1 in innate immune responses against RNA viruses. *J. Exp. Med.* **203**, 1795–1803 (2006).
- Yoneyama, M. et al. Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J. Immunol.* **175**, 2851–2858 (2005).
- Sumpter, R. Jr et al. Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *J. Virol.* **79**, 2689–2699 (2005).
- Thakur, C. S., Xu, Z., Wang, Z., Novince, Z. & Silverman, R. H. A convenient and sensitive fluorescence resonance energy transfer assay for RNase L and 2',5'-oligoadenylates. *Methods Mol. Med.* **116**, 103–113 (2005).
- Hornung, V. et al. 5'-Triphosphate RNA is the ligand for RIG-I. *Science* **314**, 994–997 (2006).
- Pichlmair, A. et al. RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* **314**, 997–1001 (2006).
- Saito, T. et al. Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. *Proc. Natl Acad. Sci. USA* **104**, 582–587 (2007).
- Malathi, K. et al. A transcriptional signaling pathway in the IFN system mediated by 2'-5'-oligoadenylate activation of RNase L. *Proc. Natl Acad. Sci. USA* **102**, 14533–14538 (2005).
- Kubota, K. et al. Identification of 2'-phosphodiesterase, which plays a role in the 2-5A system regulated by interferon. *J. Biol. Chem.* **279**, 37832–37841 (2004).
- Arimoto, K. et al. Negative regulation of the RIG-I signaling by the ubiquitin ligase RNF125. *Proc. Natl Acad. Sci. USA* **104**, 7500–7505 (2007).
- Flodstrom-Tullberg, M. et al. RNase L and double-stranded RNA-dependent protein kinase exert complementary roles in islet cell defense during coxsackievirus infection. *J. Immunol.* **174**, 1171–1177 (2005).
- Samuel, M. A. et al. PKR and RNase L contribute to protection against lethal West Nile Virus infection by controlling early viral spread in the periphery and replication in neurons. *J. Virol.* **80**, 7009–7019 (2006).
- Urisman, A. et al. Identification of a novel gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. *PLoS Pathog* **2**, e25 (2006).
- Beattie, E. et al. Reversal of the interferon-sensitive phenotype of a vaccinia virus lacking E3L by expression of the reovirus S4 gene. *J. Virol.* **69**, 499–505 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Diamond and M. Colonna (St Louis, Missouri, USA) for the *Mda5*^{-/-} MEFs, S. Akira (Osaka, Japan) for *Ips1*^{-/-} and *Rig-i*^{-/-} cells, M. David (San Diego, California, USA) for IRF3 antibodies, C.M. Rice (New York, New York, USA) for Huh7 and Huh7.5 cells, I.M. Kerr (London, UK) for EMCV, P.J. Sims (Rochester, New York, USA) for discussions, and (all from Cleveland, Ohio, USA) G. Sen for Sendai virus, J. Paranjape for cell line preparations and RNA chip analysis, B.K. Jha for RNase L, B.K. Jha, C. Thakur and Z. Novince for preparing 2-5A, and S. Shelby for technical assistance with mice. These studies were supported by grants from the NIH to R.H.S. and M.G. and by a grant from the Burroughs Wellcome Fund to M.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.H.S. (silverrr@ccf.org).

METHODS

Plasmids, reagents, cells and viruses. Plasmids were hIFN β -luciferase reporter (hIFN- β -luc)¹⁹, Renilla luciferase vector (pRL-TK) (Promega), pcDNA3-hRNaseL (human RNase L) and the same construct with R667A or H672A mutations in RNase L^{11,24}. Recombinant, purified human RNase L¹⁶, RNA FRET probe²⁰, and 2'-5'-oligoadenylates²⁴ were as described. Poly(I):poly(C) was from Amersham Biosciences (GE Healthcare). Antibodies were rabbit polyclonal antibodies against human proteins, IRF3 (a gift from M. David), RIG-I and LGP2 (M. Gale, Seattle, Washington, USA), MAVS (IPS-1) (Bethyl Labs), MDA5 (Alexis Biochemicals), STAT1 (Santa Cruz), phosphoSTAT1 (Cell Signalling Technologies) and anti- β -actin (Sigma). IFN- β was measured using mouse IFN- β ELISA kits (PBL Laboratories). Wild-type and RNaseL-deficient primary MEFs¹⁵ on C57/bl6 background were transformed with SV40 T antigen and cultured in complete RPMI medium. DU145 cells, HT1080 cells, Huh7 cells, Huh7.5 cells, RIG-I-deficient MEFs (a gift from S. Akira)³, MDA5-deficient MEFs (a gift from M. Diamond and M. Colonna)⁴ and IPS-1-deficient MEFs (a gift from S. Akira)¹⁷, and culture conditions, have been described^{19,24}. SeV (Cantell strain) from Charles River Laboratories was a gift from G. Sen and EMCV was a gift from I. M. Kerr.

Isolation and purification of RNA. Total RNA from RNase L-deficient MEFs (2.0 mg isolated using Trizol Reagent (Invitrogen)) was cleaved with RNase L (20 μ g) activated by unfractionated 2-5A (10 μ M) in the presence of a FRET RNA probe (1.0 μ g) in 500 μ l containing 25 mM Tris HCl pH 7.4, 100 mM KCl, 10 mM MgCl₂, 50 μ M ATP and 7 mM β -mercaptoethanol for 1 h at 22 °C. Control reactions lacked RNase L and/or 2-5A. RNA cleavage at 22 °C was monitored in 5- μ l samples withdrawn at different times by fluorescence with a Wallac 1420 Victor2 multilabel counter (Perkin-Elmer Life Sciences) (excitation 485 nm; emission 535 nm; 0.1 s) and in RNA chips using an Agilent Bioanalyser 2100 (Agilent Technologies). RNA size markers were from Ambion (AM7152 RNA 6000 Ladder). Small (<200-nucleotide) RNAs were isolated using *mirVana* miRNA Isolation Kit (Ambion). Small RNAs (150 μ g per reaction) were incubated without or with 30 U CIP (NEB) for 2 h at 37 °C and 10 min for 75 °C and recovered using *mirVana* miRNA Isolation Kit.

Transfections, luciferase assays, and infections of cells. Transfections of 2-5A, RNA or plasmids were done with lipofectamine 2000 (Invitrogen) whereas poly(I):poly(C) was transfected with Fugene 6 (Roche Applied Science). Cells (2×10^5 per well in six-well plates) were transiently transfected with hIFN- β -luc (1 μ g) and pRL-TK (0.1 μ g) and subjected to dual luciferase assays (Promega) in triplicate. siRNA oligonucleotides (5 nM) targeting human RIG-I, MDA5, IPS1 or LGP2 (see below), or non-specific oligonucleotides (Dharmacon), were transfected into DU145 cells with DharmaFECT 1. After 48 h, cells were transfected with luciferase-reporter plasmids and after another 24 h with 2-5A. Luciferase

activity was measured 18 h later. Alternately, 24 h after transfecting with the luciferase plasmids, cells were transfected with 2-5A or RNA for 18 h. SeV in serum-free media was added to cells in six-well plates. After 1 h, the medium was replaced with growth medium for an additional 18 h. Error bars shown are standard deviations (s.d.). Two-tailed paired Student's *t*-tests were used to calculate *P* values.

Immunoblots. Proteins in cell extracts were separated in 10% polyacrylamide/SDS gels or under native conditions through pre-cast 7.5% or 10% Tris.HCl gels (BioRad) (for RIG-I multimerization and IRF-3 dimerization assays) and transferred to nitrocellulose membrane (BioRad). Blots were incubated with various primary antibodies. Secondary goat anti-rabbit antibody tagged with horseradish peroxidase was from Cell Signalling. Immunoreactive bands were detected by enhanced chemiluminescence (ECL) (Amersham Biosciences) and subsequent exposure to X-ray film (Eastman Kodak).

Infections and 2-5A treatment of mice. Wild-type and *RnaseL*^{-/-} mice (*n* = 15 per group, 5–6 weeks old) on a C57BL/6 background were inoculated intraperitoneally with 1×10^3 pfu of EMCV or intranasally with 320 HAU of SeV. Sera were collected (at the times indicated) after infection and IFN- β levels determined by ELISA (PBL Biomedical Labs). Control mice received only PBS. For 2-5A treatment, we injected wild-type and *RnaseL*^{-/-} mice (*n* = 6 per group, 5–6 weeks old) intraperitoneally with 500 μ l of 0.6 mM unfractionated 2-5A with 80 μ l of fucose 6 for 5 h. Mock treatments were with fucose 6 alone.

Sequences of siRNA oligonucleotides. The following oligonucleotides were used (for each target mRNA, the sense sequence is given first followed by the antisense version). Human *LGP2*: GCAAUGUGGUGGUGCGUUAUU/5'-p-UUACGCACCACCACAUUGCUU; GCCAGUACCUAGAACUUAUU/5'-p-UUAAGUUCUAGGUACUGGCUU; ACAGGGAGCAGCUCACUAAUU/5'-p-UUAGUGACGUGCUCCUGUUU; CAACUUCUCGAACUACUAAUU/5'-p-AUAGUAGUUCGAGAAGUUGUU; Human *RIG-I*: GCACAGAAGUGUAU-AUUGGUU/5'-p-CCAAUAUACACUUCUGUGCUU; CCACAACACUAGU-AAACAAUU/5'-p-UUGUUUACUAGUGUUGUGGUU; CGGAUUAGCGAC-AAAUUUUU/5'-p-UAAUUUUGUGCGUAAUCCGUU; UCGAUGAGAU-UGAGCAAGAUU/5'-p-UCUUGCUCAAUUCUACUGAUU; Human *MDA5*: GAAUAACCCAUCACUAAUAAUU/5'-p-UAUUAGUGAUGGGUUAUUCUU; GCACGAGGAUAAUUCUUUUUU/5'-p-UAAAGAUUAUUCUUGUGCUU; UGACACAAUUCGAAUGAUUU/5'-p-UAUCAUUCGAAUUGUGUCAUU; CAAUGAGGCCCCUACAAUUUU/5'-p-AAUUUGUAGGGCCUACUUGUU; Human *IPS-1*: AAGUAUAUCUGCCGCAUUUU/5'-p-AAUUGCGGCAGAUUACUUUU; CAUCCAAAGUGCCUACUAGUU/5'-p-CUAGUAGGCAC-UUUUGGAUGUU; GCAAUGUGGAUGUUGUAGAUU/5'-p-UCUACAACAU-CCACAUUGCUU; CAUCCAAUUGCCCAUCAAUU/5'-p-UUGAUGGGC-AAUUUGGAUGUU.

LETTERS

Break-induced replication and telomerase-independent telomere maintenance require Pol32

John R. Lydeard¹, Suvi Jain¹, Miyuki Yamaguchi¹ & James E. Haber¹

Break-induced replication (BIR) is an efficient homologous recombination process to initiate DNA replication when only one end of a chromosome double-strand break shares homology with a template^{1–5}. BIR is thought to re-establish replication at stalled and broken replication forks and to act at eroding telomeres in cells that lack telomerase in pathways known as ‘alternative lengthening of telomeres’ (reviewed in refs 2, 6). Here we show that, in haploid budding yeast, Rad51-dependent BIR induced by HO endonuclease requires the lagging strand DNA Pol α -primase complex as well as Pol δ to initiate new DNA synthesis. Pol δ is not required for the initial primer extension step of BIR but is required to complete 30 kb of new DNA synthesis. Initiation of BIR also requires the nonessential DNA Pol δ subunit Pol32 primarily through its interaction with another Pol δ subunit, Pol31. HO-induced gene conversion, in which both ends of a double-strand break engage in homologous recombination, does not require Pol32. Pol32 is also required for the recovery of both Rad51-dependent and Rad51-independent survivors in yeast strains lacking telomerase. These results strongly suggest that both types of telomere maintenance pathways occur by recombination-dependent DNA replication. Thus Pol32, dispensable for replication and for gene conversion, is uniquely required for BIR; this finding provides an opening into understanding how DNA replication re-start mechanisms operate in eukaryotes. We also note that Pol32 homologues have been identified both in fission yeast and in metazoans where telomerase-independent survivors with alternative telomere maintenance have also been identified^{2,6,7}.

When a double-strand break (DSB) occurs in the middle of a chromosome, it can be efficiently repaired by gene conversion, using homologous sequences on a sister chromatid, a homologous chromosome, or at an ectopic location^{1,8,9}. However, if only one end of a DSB shares homology with other sequences in the genome, cells rely on recombination-dependent DNA replication (known as break-induced replication, BIR) to restore genome integrity^{2–5} (see also Supplementary Fig. 1). Such events are believed to occur at stalled replication forks, where endonuclease cleavage produces one partially replicated chromatid and an intact chromatid^{10,11}. BIR can re-establish a unidirectional replication fork, which can proceed to the end of the template chromatid or until it meets a converging replication fork.

BIR also could explain how rare survivors arise in *Saccharomyces cerevisiae* cells lacking telomerase, in which chromosome ends become eroded and recombinogenic¹². Virtually all telomerase-deficient survivors require the key homologous recombination protein Rad52, but there are two distinct survivor pathways that differ in their genetic and substrate requirements^{13,14}. Type I Rad51-dependent survivors result from recombination between subtelomeric repeat sequences, whereas Type II Rad51-independent—but Rad50- and Rad59-dependent—survivors exhibit recombination between imperfect nucleotide

TG_{1–3} sequences of the telomeres themselves. It is believed that these events must involve DNA replication—possibly rolling circle replication of an excised circular template²—but there has not yet been any direct demonstration of the involvement of leading- and lagging-strand DNA polymerases in these repair events.

We developed a haploid system to study BIR in budding yeast, where a galactose-inducible HO endonuclease creates a DSB adjacent to the 5' part of the *CAN1* gene (designated CA) in a nonessential terminal region of chromosome V-left (Chr. V-L). The DSB shares homology, only on the centromere-proximal side, with a 3' segment of the *CAN1* gene (called AN1) that is inserted on another chromosome (Fig. 1a). Successful recombination results in a non-reciprocal translocation, restoring an intact *CAN1* gene that renders the cells canavanine sensitive, but leads to the loss of the hygromycin-resistance *HPH-MX* marker distal to the DSB. Interchromosomal Rad51- and Rad52-dependent recombination yields BIR products, the frequency of which increases from 21% to 31% as the length of shared homology on Chr. XI increases from 1,157 to 2,977 base pairs (bp) (Fig. 1b and Supplementary Table 1). BIR was more efficient in

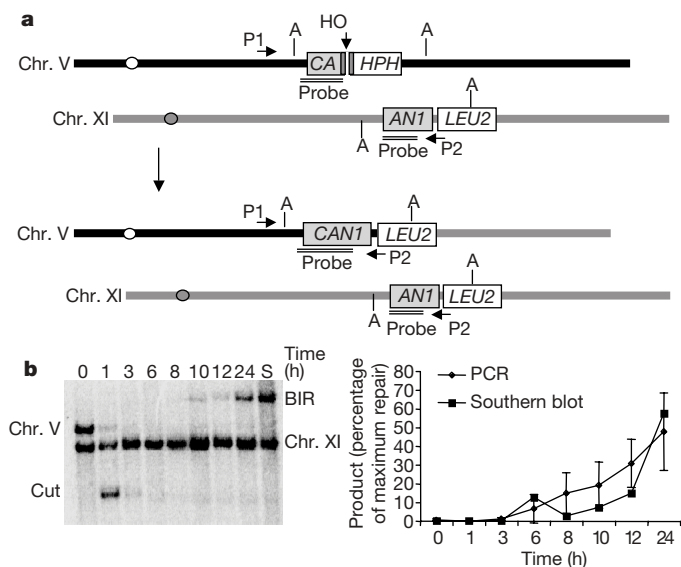


Figure 1 | Experimental system to study BIR and kinetics of DSB repair. **a**, An *HPH-MX* marked HO endonuclease cut site (grey bars) integrated into Chr. V-L deletes the 3' end of *CAN1*. The *LEU2*-marked *AN1* donor sharing 1,157 bp of homology with *CAN1* is integrated into Chr. XI. **b**, PCR with primers P1 and P2 monitors initiation of new DNA synthesis. Southern blot analysis of *Ava*I-digested DNA (marked by A) monitors completion of BIR. Lane S contains DNA from a colony in which BIR occurred. Kinetics of repair is shown for PCR and Southern blot assays of BIR induced in cycling cells. Error bars represent s.d.

¹MS029 Rosenstiel Centre, Brandeis University, Waltham, Massachusetts 02454-9110, USA.

intrachromosomal events in which the *ANI* sequences were inserted on the opposite arm of Chr. V, in opposite orientation (Supplementary Table 1). There may be an effect of the length of the template that BIR must copy, because the efficiency of using the same size 1,157-bp *ANI* homology was reduced for inserts 50 kb from the Chr. XI telomere or 200 kb from the Chr. V-R telomere. However, a more systematic study will be required to determine if the effects reflect distance or some other sequence context. Deleting *RRM3*, important in allowing DNA replication across difficult-to-replicate regions^{15,16}, did not affect repair with the *ANI* template 200 kb from the telomere (Supplementary Table 1).

BIR presumably involves recombination-dependent establishment of a unidirectional DNA replication fork involving leading and lagging strand synthesis. To study essential DNA replication functions we used temperature-sensitive mutations. Cells were arrested in nocodazole to ensure that they had completed DNA replication so that DNA replication components were not sequestered, a problem we previously encountered in showing that Pol α -primase is not required for gene conversion^{17,18}. Gene conversion probably arises from synthesis-dependent strand annealing, in which the 3' ends of the DSB are extended after strand invasion of the template, but then dissociate and anneal¹⁹. Here we show that temperature-sensitive mutations of primase (*pri2-1*) and Pol α (*pol1-17*) prevent the completion of HO-endonuclease-induced BIR.

After wild-type and mutant cells were grown at the permissive temperature of 25 °C and arrested with nocodazole, cultures were split and either maintained at the permissive temperature or were shifted to the restrictive temperature of 37 °C and, after 3 h, galactose was added to the medium to induce expression of HO endonuclease. As shown in Fig. 2 and Supplementary Fig. 2, BIR (measured by a PCR-based assay that detects new DNA synthesis extended from the 3' end of the invading strand) was almost completely eliminated in the temperature sensitive primase strain at 37 °C ($P < 0.05$ at 10 h and later). Even at the permissive temperature, viability of *pri2-1* cells was only 12% ($P = 0.002$) after inducing a DSB, compared with a wild-type viability of 21%. A significant defect was also observed in *pol1-17* cells at 37 °C ($P < 0.05$ at 8 h and later). These data demonstrate for the first time that BIR requires the lagging-strand Pol α -primase complex. We note that BIR is significantly slower in nocodazole-arrested cells than in cycling cells, possibly because some components of DNA replication are not present as abundantly in G2-arrested cells.

We next asked if the other essential DNA polymerases, Pol δ and Pol ϵ , are required for BIR. We created derivatives of strain JRL092

carrying either a temperature-sensitive allele of the catalytic subunit of Pol δ , (*pol3-14*)²⁰ or a temperature-sensitive allele of the catalytic subunit of Pol ϵ (*pol2-11*)²¹, both of which prevent DNA replication at their restrictive temperature. BIR was induced in nocodazole-arrested cells and repair was monitored by the PCR assay described above. Primer extension was observed in both *pol3-14* and *pol2-11* cells at their permissive temperature (25 °C). There was little repair in *pol3-14* cells at the restrictive temperature, whereas the efficiency of primer extension in *pol2-11* cells was unaffected (Fig. 2a). Thus, whereas the Pol α -primase complex and Pol δ are required for the initiation of new BIR replication, Pol ϵ is not.

Although Pol ϵ is not required for the initial primer extension, it is required to complete 30 kb of DNA replication in BIR, as seen on a pulsed-field gel of intact chromosomes (Fig. 2b). At the restrictive temperature of 37 °C, a *pol2-11* mutant shows a fourfold reduction in the level of BIR product compared with the wild-type strain at the same temperature. Thus, initiation of primer extension requires Pol δ and Pol α -primase, but Pol ϵ must play an important part in the processive elongation of the newly synthesized DNA to the end of the template. The different roles of Pol δ and Pol ϵ in BIR may explain some striking results²² of frequent template-switching events during BIR, with switches confined to the first few kilobases of new synthesis. A transition from the sole use of Pol δ as BIR is initiated to a more important role for Pol ϵ during the elongation phase of BIR might account for these observations.

The nonessential DNA Pol δ subunit Pol32 has been implicated in post-replicative repair and in mutagenesis associated with bypass of DNA damage^{23–26}. Here we show that Pol32 is essential for BIR, whereas deletion of either of two nonessential subunits of DNA polymerase ϵ has no effect (Fig. 3a). Southern blot and primer-extension analysis confirmed that there was no product formation (Fig. 3b, Supplementary Fig. 2b and data not shown). In contrast, HO-endonuclease-induced gene conversion arising by ectopic recombination is only slightly affected by the absence of Pol32 (Fig. 4a). We examined gene conversion in which an HO-endonuclease-induced DSB is created within a 117-bp cleavage site inserted into the *leu2* gene²⁷, which was inserted at the *CAN1* gene on chromosome V. The DSB in strain YSJ119 could be repaired efficiently by gene conversion, whereas, in strain YSJ143, repair proceeds by BIR in about 12% of cells. In both cases, a nonhomologous segment of at least 47 bp of the HO endonuclease cleavage site must be removed before new DNA synthesis can be extended from the 3' end after strand invasion. Deletion of *POL32* caused a 25% reduction in gene conversion efficiency, but it nearly completely eliminated BIR,

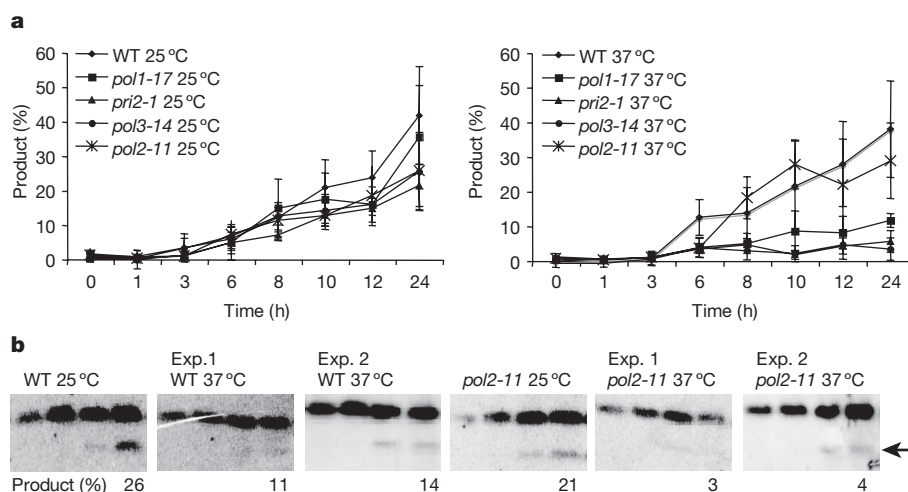


Figure 2 | The Pol α -primase complex, Pol δ and Pol ϵ are required for BIR. **a**, Appearance of BIR repair product, as monitored by PCR, in wild-type (WT), *pol1-17*, *pri2-1*, *pol3-14* and *pol2-11* cells arrested in nocodazole at either 25 °C or 37 °C. Data are the mean \pm s.d. **b**, Completion of BIR as

monitored by pulsed-field gel electrophoresis followed by hybridization with a probe for *MCH2*, which normally lies 6 kb from the telomere on Chr. XI. The BIR product, a non-reciprocal translocation between Chr. V and the terminal 30 kb of Chr. XI (Fig. 1a), is shown by an arrow. Exp., experiment.

showing a reduction of at least 20-fold (Fig. 4a, Supplementary Table 2). Deleting *POL32* also had very little effect on the efficiency or kinetics of repair of another gene conversion event, HO-endonuclease-induced *MAT* switching (Supplementary Fig. 3). We have previously presented data showing that gene conversion events arise primarily by synthesis-dependent strand annealing, in which only DNA synthesis primed from the 3' end of the invading strand is required¹⁸. These data suggest that Pol32 plays a crucial role in recombination-dependent establishment of a repair replication fork during BIR and is not needed for all types of recombination-associated repair DNA synthesis.

A series of domain mutations of Pol32 that eliminate protein-protein interactions between Pol32 and another Pol δ subunit, Pol31, or between Pol α and proliferating cell nuclear antigen (PCNA) have been created²⁸. We introduced a series of centromeric plasmids carrying wild-type *POL32* or *pol32* mutations eliminating one of the interactions. Complementation of the *pol32* Δ phenotype

was found with the *POL32* plasmid, whereas the *pol32-2* mutation that eliminated Pol32–Pol31 binding was as defective as *pol32* Δ (Fig. 3c). Deleting the PCNA-interacting domain (*pol32-10*) resulted in a statistically significant decrease in BIR, whereas a small deletion disrupting the Pol32–Pol α binding (*pol32-8*) had no significant effect on BIR (Fig. 3c). However, larger deletions in this domain (*pol32-11*, *pol32-26* and *pol32-27*) reduced the percentage of cells completing BIR (Fig. 3c). These results show that Pol32's role in BIR depends primarily on its interaction with Pol31, but that its interaction with both PCNA and the Pol α -primase complex are also probably important. We also note that deleting two nonessential subunits of Pol ϵ , Dpb3 and Dpb4, has no effect on BIR (Fig. 3a and Supplementary Table 1).

It has been presumed that the Rad51-dependent and -independent 'alternative lengthening of telomeres' pathways in budding yeast result from recombination-dependent DNA replication². To test this idea directly we created haploid strains lacking both the gene encoding telomerase RNA *TLC1* and *POL32*. Meiotic segregants were propagated by dilution of overnight cultures sequentially through 8 days, as previously described¹³. Wild-type and *pol32* Δ *TLC1* cells maintain their viability through this regime, whereas *tlc1* Δ shows a progressive loss of viability followed by the appearance of survivors that are able to grow to the same extent as wild type (Fig. 4b). In contrast, *tlc1* Δ *pol32* Δ cells fail to yield any survivors after 9 days and resemble *rad52* Δ *tlc1* Δ cells in their loss of viability and failure to yield any survivors when plated after 8 days of culture¹³. These results demonstrate that both Rad51-dependent and Rad51-independent survivor pathways in *tlc1* Δ cells rely on Pol32, because cells

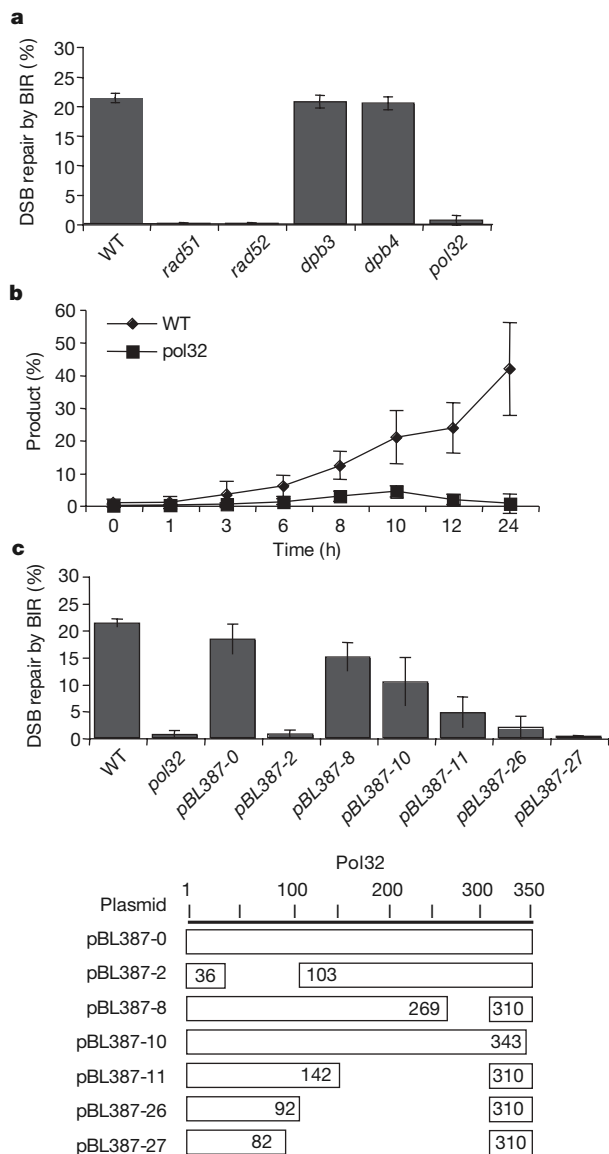


Figure 3 | *POL32* is required for BIR. **a**, Efficiency of BIR in *rad51* Δ , *rad52* Δ , *pol32* Δ , *dpb3* Δ , and *dpb4* Δ cells, as measured by viability following a DSB. **b**, *POL32* is required for primer extension of BIR. Appearance of BIR repair product, as monitored by PCR, is shown for wild-type and *pol32* Δ cells. **c**, The effect of different *pol32* alleles on BIR. The positions of the deletions in each allele were determined by Johansson *et al.*²⁸ (figure adapted with permission). The last amino acid to the left and the first amino acid to the right of each deletion is shown. Error bars represent s.d.

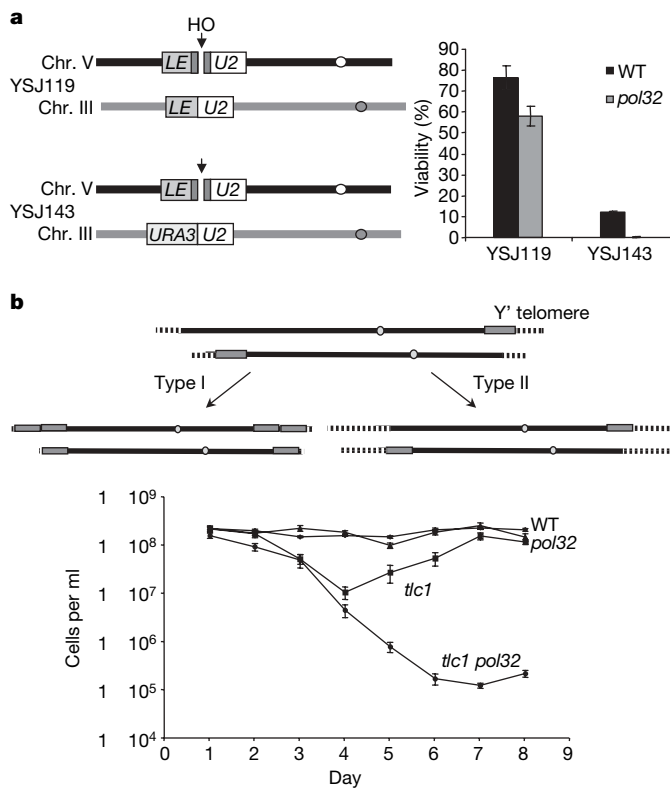


Figure 4 | Effect of *POL32* on gene conversion and on telomere maintenance. **a**, *POL32* is required for BIR but not gene conversion. A DSB within *leu2* on Chr. V-L can be repaired by gene conversion with a *LEU2* donor on Chr. III (YSJ119) or by BIR (YSJ143) sharing homology only to *U2*. A ≥ 47 -bp non-homologous segment (dark vertical bar) must be removed before new DNA synthesis can be initiated. **b**, *POL32* is required for both Type I and Type II telomere maintenance in the absence of telomerase. Meiotic segregants lacking *TLC1*, and/or *POL32* were grown to saturation, diluted to 5×10^5 cells ml⁻¹ and transferred to fresh medium; cell density was measured 24 h later¹³. Error bars represent s.d.

individually lacking Rad51 or either Rad50 or Rad59 (which are required for Rad51-independent survivors) still yield Type I or Type II survivors, respectively^{13,14,29,30}. Taken together, these results demonstrate that recombination-dependent survivor pathways in budding yeast are all replication-dependent and presumably involve both leading- and lagging-strand synthesis. We suggest that Pol32 is essential in establishing a full replication fork by recombination, in the absence of an origin of replication. Alternative lengthening of telomere (ALT) pathways that are seen in many transformed mammalian cell lines and some human cancers may also use similar BIR pathways.

METHODS SUMMARY

Yeast strains. The strains used and their derivations are described in detail in the Methods.

Viability measurements. Logarithmically growing cells were grown in YEP-lactate or the appropriate drop-out media +2% Raffinose, plated onto either YEPD or YEP-Gal, and grown into colonies. Cell viability after HO-endonuclease-induction was derived by dividing the number of colony forming units (CFU) on YEP-Gal by that on YEPD. A more detailed description is found in the Methods. Survival in strains lacking the *TLC1* gene encoding telomerase RNA was assessed as described¹³.

HO endonuclease induction and measurement of kinetics of DSB repair. Strains were grown in YP-lactate (1% Bacto yeast extract, 2% Bacto peptone, 3% lactic acid, pH 5.5) and time course experiments were performed as previously described¹⁸.

DNA analysis. DNA analysis to monitor the kinetics of repair by Southern blot are described in the Methods. To assay for the completion of BIR, DNA was analysed by pulsed-field clamped homogeneous electric field (CHEF) gel electrophoresis followed by Southern blot. Membranes were then probed with *MCH2* fragment generated by PCR. Parameters of CHEF gels are described in the Methods. For PCR analysis, equal amounts of purified genomic DNA were diluted and amplified within the linear range with primers specific to the HO-endonuclease-cut chromosome and the donor region. All PCR products were subjected to electrophoresis in 1% agarose, stained with ethidium bromide, and quantified using Bio-Rad Quantity One software. The percentage of BIR product formed was determined by dividing the BIR product signal to that amplified from an independent locus from the same input and under the same conditions and then normalized to the ratio from a surviving colony and plotted against time. A more detailed explanation of PCR analysis is described in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 May; accepted 25 June 2007.

Published online 1 August 2007.

- Haber, J. E. Transpositions and translocations induced by site-specific double-strand breaks in budding yeast. *DNA Repair* **5**, 998–1009 (2006).
- McEachern, M. J. & Haber, J. E. Break-induced replication and recombinational telomere elongation in yeast. *Annu. Rev. Biochem.* **75**, 111–135 (2006).
- Morrow, D. M., Connolly, C. & Hieter, P. "Break copy" duplication: a model for chromosome fragment formation in *Saccharomyces cerevisiae*. *Genetics* **147**, 371–382 (1997).
- Davis, A. P. & Symington, L. S. *RAD51*-dependent break-induced replication in yeast. *Mol. Cell. Biol.* **24**, 2344–2351 (2004).
- Malkova, A., Naylor, M., Yamaguchi, M., Ira, G. & Haber, J. E. *RAD51*-dependent break-induced replication differs in kinetics and checkpoint responses from *RAD51*-mediated gene conversion. *Mol. Cell. Biol.* **25**, 933–944 (2005).
- Henson, J. D., Neumann, A. A., Yeager, T. R. & Reddel, R. R. Alternative lengthening of telomeres in mammalian cells. *Oncogene* **21**, 598–610 (2002).
- Nakamura, T. M., Cooper, J. P. & Cech, T. R. Two modes of survival of fission yeast without telomerase. *Science* **282**, 493–496 (1998).
- Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).
- Symington, L. S. Role of *RAD52* epistasis group genes in homologous recombination and double-strand break repair. *Microbiol. Mol. Biol. Rev.* **66**, 630–670 (2002).
- Michel, B. Replication fork arrest and DNA recombination. *Trends Biochem. Sci.* **25**, 173–178 (2000).

- Haber, J. E. DNA recombination: the replication connection. *Trends Biochem. Sci.* **24**, 271–275 (1999).
- Lundblad, V. & Blackburn, E. H. An alternative pathway for yeast telomere maintenance rescues *est1*-senescence. *Cell* **73**, 347–360 (1993).
- Le, S., Moore, J. K., Haber, J. E. & Greider, C. *RAD50* and *RAD51* define two different pathways that collaborate to maintain telomeres in the absence of telomerase. *Genetics* **152**, 143–152 (1999).
- Teng, S., Chang, J., McCowan, B. & Zakian, V. A. Telomerase-independent lengthening of yeast telomeres occurs by an abrupt Rad50p-dependent, Rif-inhibited recombinational process. *Mol. Cell* **6**, 947–952 (2000).
- Torres, J. Z., Schnakenberg, S. L. & Zakian, V. A. *Saccharomyces cerevisiae* Rrm3p DNA helicase promotes genome integrity by preventing replication fork stalling: viability of *rrm3* cells requires the intra-S-phase checkpoint and fork restart activities. *Mol. Cell. Biol.* **24**, 3198–3212 (2004).
- Torres, J. Z., Bessler, J. B. & Zakian, V. A. Local chromatin structure at the ribosomal DNA causes replication fork pausing and genome instability in the absence of the *S. cerevisiae* DNA helicase Rrm3p. *Genes Dev.* **18**, 498–503 (2004).
- Holmes, A. & Haber, J. E. Double-strand break repair in yeast requires both leading and lagging strand DNA polymerases. *Cell* **96**, 415–424 (1999).
- Wang, X. *et al.* Role of DNA replication proteins in double-strand break-induced recombination in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **24**, 6891–6899 (2004).
- Ira, G., Satory, D. & Haber, J. E. Conservative inheritance of newly synthesized DNA in double-strand break-induced gene conversion. *Mol. Cell. Biol.* **26**, 9424–9429 (2006).
- Giot, L., Chanet, R., Simon, M., Facca, C. & Faye, G. Involvement of the yeast DNA polymerase delta in DNA repair *in vivo*. *Genetics* **146**, 1239–1251 (1997).
- Budd, M. E. & Campbell, J. L. DNA polymerases δ and ϵ are required for chromosomal replication in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**, 496–505 (1993).
- Smith, C. E., Llorente, B. & Symington, L. S. Template switching during break-induced replication. *Nature* **447**, 102–105 (2007).
- Johansson, E., Majka, J. & Burgers, P. M. Structure of DNA polymerase δ from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**, 43824–43828 (2001).
- Corrette-Bennett, S. E., Borgeson, C., Sommer, D., Burgers, P. M. & Lahue, R. S. DNA polymerase δ , RFC and PCNA are required for repair synthesis of large looped heteroduplexes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **32**, 6268–6275 (2004).
- Huang, M. E., de Calignon, A., Nicolas, A. & Galibert, F. POL32, a subunit of the *Saccharomyces cerevisiae* DNA polymerase δ , defines a link between DNA replication and the mutagenic bypass repair pathway. *Curr. Genet.* **38**, 178–187 (2000).
- Huang, M. E., Rio, A. G., Galibert, M. D. & Galibert, F. Pol32, a subunit of *Saccharomyces cerevisiae* DNA polymerase δ , suppresses genomic deletions and is involved in the mutagenic bypass pathway. *Genetics* **160**, 1409–1422 (2002).
- Vaze, M. *et al.* Recovery from checkpoint-mediated arrest after repair of a double-strand break requires srs2 helicase. *Mol. Cell* **10**, 373–385 (2002).
- Johansson, E., Garg, P. & Burgers, P. M. The Pol32 subunit of DNA polymerase δ contains separable domains for processive replication and proliferating cell nuclear antigen (PCNA) binding. *J. Biol. Chem.* **279**, 1907–1915 (2004).
- Chen, Q., Ijima, A. & Greider, C. W. Two survivor pathways that allow growth in the absence of telomerase are generated by distinct telomere recombination events. *Mol. Cell. Biol.* **21**, 1819–1827 (2001).
- Teng, S. C. & Zakian, V. A. Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 8083–8093 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to P. Burgers for his generosity in providing Pol32 mutant plasmids and for comments, to M. Budd and J. Campbell for plasmids, and to members of the Haber laboratory for their comments. This work was supported by the NIH. J.R.L. was supported by an NIH Genetics Training Grant.

Author Contributions J.R.L., S.J., M.Y. and J.E.H. designed experiments. J.R.L. carried out experiments and analysed BIR in the *CAN1* system, S.J. carried out and analysed experiments in the *LEU2* system and M.Y. performed experiments and analysed the MAT switching system and survivors in the absence of *TLC1*. J.R.L. and J.E.H. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.E.H. (haber@brandeis.edu).

METHODS

Strains. Strains used to study BIR are derivatives of the S288c parent strain CL11-7 (*hoA mat::hisG hmlΔ::hisG HMRA-stk ura3Δ851 trp1Δ63 leu2Δ::KAN ade3::GAL10::HO*). To introduce a HO cut site into chromosome V (Chr. V) while deleting a 3' fragment of the *CAN1* gene, an *HPH-MX* (hygromycin-resistance) marker carrying a 117-bp HO endonuclease cut site (HOcs) from *MATa* (constructed by M. Vaze) was amplified by PCR from a pMBV40 plasmid DNA with mixed oligonucleotides containing homology to the template DNA and 50 bp of homology to the target region. The resulting strain has a HOcs-*HPH* cassette inserted between nucleotides 29,176–32,020 of Chr. V, resulting in a truncation of the *CAN1* ORF at nucleotide 1,146 to create strain JRL016 (CL11-7 *can1,1-1446::HOcs::HPH*). The same approach using different primer sets was used to construct strain JRL015 (CL11-7 *can1,1-577A539::HOcs::HPH*) in which the HOcs-*HPH* cassette is inserted between nucleotides 29,176–32,927 of Chr. V, resulting in a truncation of the *CAN1* ORF at nucleotide 539. To introduce a donor into these strains a PCR fragment was amplified from plasmid pJRL8, in which the *LEU2* and *CAN1* genes are adjacent to each other in opposite orientations and transformed into the indicated parent strain. The PCR fragment was generated with mixed oligos containing homology to either *LEU2* or *CAN1* and 50 bp of homology to the target region. Different primer sets were used to amplify either a truncated *LEU2::can1Δ289* (truncated at nucleotide 289 of the *CAN1* ORF) or full-length *LEU2::CAN1* fragment into different varied loci. For strains JRL092 (JRL016 *yk1215c::LEU2::can1Δ289*) and JRL089 (JRL016 *yk1215c::LEU2::CAN1*) the donor was inserted between nucleotides 27,175–28,725 on Chr. XI, whereas the donor was integrated between nucleotides 48,314–48,900 on Chr. XI in strain JRL070 (JRL016 *yk1207w::LEU2::can1Δ289*). For strains JRL040 (JRL015 *fau1::LEU2::can1Δ289*) and JRL084 (JRL016 *fau1::LEU2::can1Δ289*), the donor was inserted between nucleotides 554,363–555,900 on the right arm of Chr. V, whereas the donor was inserted between nucleotides 387,044–386,462 of Chr. V in strain JRL062 (JRL016 *lsm4::LEU2::can1Δ289*). All Chr. V donors were oriented opposite to the native *CAN1* ORF and on the opposite side of the centromere. All strains were confirmed by PCR and Southern blot analysis. The *pol1-17*, *pri2-1*, and *pol3-14* conditional alleles were introduced into JRL092 by integration and excision as previously described and confirmed by sequencing and temperature sensitivity¹⁷. The *pol2-11* conditional allele was integrated into JRL092 as previously described and confirmed by sequencing and temperature sensitivity²¹. Unless otherwise stated, mutant strains used to study BIR are isogenic to JRL092 and were created by standard gene disruption methods.

Strains used to study HO-endonuclease-induced *MAT* switching were derived from JKM161 (*HMLα MATa hmrA::ADE1 ura3-52 leu2-3,112 trp1::hisG lys5 ade1-100 ade3::GAL::HO*)³¹. Strains used to compare BIR versus gene conversion were derived from strain YFP17 (*hoA hmlA::ADE1 mataA::hisG hmrA::ADE1 leu2::HOcs ade3::GAL::HO ade1 lys5 ura3-52*)²⁷. To introduce the *leu2::HOcs* into the *CAN1* gene on Chr. V, the endogenous *leu2::HOcs* was first replaced with the *KAN-MX* cassette at nucleotides 90,876–92,652 on Chr. III to generate strain YSJ85 (YFP17 *leu2::KAN*). Next, the *leu2::HOcs* containing the 117 bp HOcs integrated at nucleotide 91,570 (Asp718) on Chr. III was amplified by PCR from YFP17 at Chr. III 90,863–92,620 and inserted into the *CAN1* locus on Chr. V between nucleotides 31,786–33,558 by transformation to create strain YSJ100 (YSJ85 *can1::leu2::HOcs*). To generate an ectopic gene conversion system, the *LEU2* donor containing 683 bp and 1321 bp of homology to the left and right sides of the HO endonuclease cut site, respectively (Chr. III coordinates 90,887–92,891), was inserted at position 41,400 on the left arm of Chr. III by transformation to obtain strain YSJ119 (YSJ100 *YCL048w-A::LEU2*). To generate a congenic BIR system, a *URA3*-marked fragment containing *LEU2* sequence with coordinates (91,570–92,891) amplified by PCR from plasmid pSJ14 was inserted at position 41,400 on Chr. V to delete the 683 bp of homology on the left side of the break but retain the 1,321 bp of homology to the right side of the HO

endonuclease cut site. YSJ144 (YSJ119 *pol32Δ*) and YSJ150 (YSJ143 *pol32Δ*) were constructed by standard gene disruption methods.

Strain UCC3535 (*MATa/MATα*, homozygous for *ura3-52 lys2-80 ade2-101 his3-Δ200 trp1-Δ1 leu2-Δ1 DIA5-1* and heterozygous for *TLCl/tlc1Δ::LEU2*) was provided by M. Wellinger. It was transformed with *pol32Δ::KAN*. The diploid (MY067) was sporulated and freshly dissected to obtain meiotic segregants to analyse the effect of Pol32 on the formation of survivors in *tlc1Δ* strains.

Viability measurements. Logarithmically growing cells grown in YEP-lactate or the appropriate drop-out media +2% raffinose were plated on either YEPD or YEP-Gal and grown into colonies. Colonies were counted and were then replica plated onto plates containing either canavanine or hygromycin. Cell viability after HO-endonuclease-induction was derived by dividing the number of CFUs on YEP-Gal by that on YEPD. In JRL-derived strains the percentage of cells repairing by BIR was determined by the number of *Can^s, Hph^s* YEP-Gal CFU to the number of *Can^r, Hph^r* YEPD CFU. Approximately 2% of all cells, independent of genetic background tested, remained *Can^r, Hph^r* following growth on YEP-Gal, owing to repair by gene conversion using *HMRA-stk* as a donor (confirmed by sequencing). When *HMRA-stk* was deleted in JRL092, the *Can^r, Hph^r* survivors were eliminated and the percentage of repair by BIR was only moderately affected, with an increase from 21% to 23%. Experiments were performed at least 3 times for each strain. To determine the statistical significance of the *Pol32* alleles to the wild-type plasmid (pBL387-0) on BIR the Student's *t*-test was used (paired, two-tailed, *n* = 4 for all strains except for *pol32Δ* pBL387-8, for which *n* = 6, and *pol32Δ* pBL387-10, for which *n* = 5).

DNA analysis. For Southern blot analysis of BIR, purified genomic DNA was digested with *AvaI*, separated on a 0.8% Agarose gel, and probed with a ³²P-labelled *CAN1* fragment generated by PCR. The Southern blots were scanned by PhosphorImager, and the repair efficiency was calculated by comparing the ratio of the BIR product fragment to the corresponding donor fragment, normalized to the ratio from a surviving colony and plotted against time. To analyse the kinetics of gene conversion, Southern blot analysis was performed as previously described³².

To assay for the completion of BIR, DNA was analysed by CHEF gel electrophoresis followed by Southern blot. DNA was first embedded in agarose plugs as described by the manufacturer (Bio-Rad) and separated at 200 V in 0.25× Tris-borate-EDTA at 14 °C for 40 h (initial time = 5 s, final time = 100 s) using a CHEF-DR II Pulsed-Field Electrophoresis system (Bio-Rad). Gels were stained with ethidium bromide and transferred to a nylon membrane in 0.4 M NaOH and 1.5 M NaCl using a Vacuum blotter from Appligene. Membranes were then probed with *MCH2* fragment generated by PCR.

For PCR analysis, equal amounts of purified genomic DNA were diluted, and PCR was carried out using buffer F from Epicentre and Taq polymerase (Promega). All PCR of JRL092-derived strain reactions were amplified within the linear range and early BIR products were identified with primers P1 (specific to chromosome 5) and P2 (specific to the donor region on chromosome 11). All PCR product was subjected to electrophoresis in 1% agarose, stained with ethidium bromide, and quantified using Bio-Rad Quantity One software. The percentage of BIR product formed was determined by dividing the BIR product signal to that amplified from an independent locus (*FLO9*) on chromosome 1 from the same input and under the same conditions and then normalized to the ratio from a surviving colony and plotted against time. At least three PCR reactions were performed for each strain. To determine the statistical significance of the PCR signal between strains at their permissive and restrictive temperatures, the Student's *t*-test was used (paired, two-tailed, *n* = 4 for all strains except for wild-type and *pol1-17*, for which *n* = 6).

31. Sugawara, N., Wang, X. & Haber, J. E. *In vivo* roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. *Mol. Cell* 12, 209–219 (2003).
32. Sugawara, N. & Haber, J. E. Repair of DNA double strand breaks: *in vivo* biochemistry. *Methods Enzymol.* 408, 416–429 (2006).

LETTERS

A transglutaminase homologue as a condensation catalyst in antibiotic assembly lines

Pascal D. Fortin^{1*}, Christopher T. Walsh¹ & Nathan A. Magarvey^{1*}

The unrelenting emergence of antibiotic-resistant bacterial pathogens demands the investigation of antibiotics with new modes of action. The pseudopeptide antibiotic andrimid is a nanomolar inhibitor of the bacterial acetyl-CoA carboxylase that catalyses the first committed step in prokaryotic fatty acid biosynthesis¹. Recently, the andrimid (*adm*) biosynthetic gene cluster was isolated and heterologously expressed in *Escherichia coli*². This establishes a heterologous biological host in which to rapidly probe features of andrimid formation and to use biosynthetic engineering to make unnatural variants of this important and promising new class of antibiotics. Bioinformatic analysis of the *adm* cluster revealed a dissociated biosynthetic assembly system lacking canonical amide synthases between the first three carrier protein domains. Here we report that AdmF, a transglutaminase (TGase) homologue, catalyses the formation of the first amide bond, an *N*-acyl- β -peptide link, in andrimid biosynthesis. Hence, AdmF is a newly discovered biosynthetic enzyme that acts as a stand-alone amide synthase between protein-bound, thiotemplated substrates in an antibiotic enzymatic assembly line. TGases (enzyme class (EC) 2.3.2.13) normally catalyse the cross-linking of (poly)peptides by creating isopeptidic bonds between the γ -carboxamide group of a glutamine side chain of one protein and various amine donors, including lysine side chains³. To the best of our knowledge, the present study constitutes the first report of a TGase-like enzyme recruited for the assembly of an antibiotic. Moreover, genome mining using the AdmF sequence yielded additional

TGases in unassigned natural product biosynthetic pathways. With many more microbial genomes being sequenced, such a strategy could potentially unearth biosynthetic pathways producing new classes of antibiotics.

Andrimid, isolated from different bacteria around the globe^{2,4–6}, is a hybrid nonribosomal peptide (NRP)-polyketide antibiotic made of a central β -amino acid/ α -amino acid (β -phenylalanine-L-valine) dipeptide core that is amino-capped by a polyunsaturated fatty acid (octa-2,4,6-trienoic acid) and carboxy-capped by a methylsuccinimide moiety (Fig. 1a, and Supplementary Fig. 1)⁶. Like other hybrid polyketide and NRP natural products, andrimid is in part synthesized in the active sites of the multi-functional enzymes polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs), using a thiotemplate logic. The known PKS and NRPS thiotemplate assembly logic involves the sequestration of monomeric building blocks as covalently tethered units by a thioester linkage to the phosphopantetheinyl prosthetic groups of carrier-protein domains, otherwise known as thiolation (T) domains. Tethered acyl or amino-acyl chains serve as substrates in subsequent elongation reactions using one of two condensation catalysts; ketosynthase domains (KS) in PKSs or condensation (C) domains in NRPSs⁷.

Many of the antibiotics discovered are naturally produced by multi-functional enzyme assembly lines, which, in essence, consist of composites of a number of catalytic and structural domains fused in *cis*. In many cases the enzymatic steps and the sequence leading to the production of an antibiotic can be inferred by the relative order of

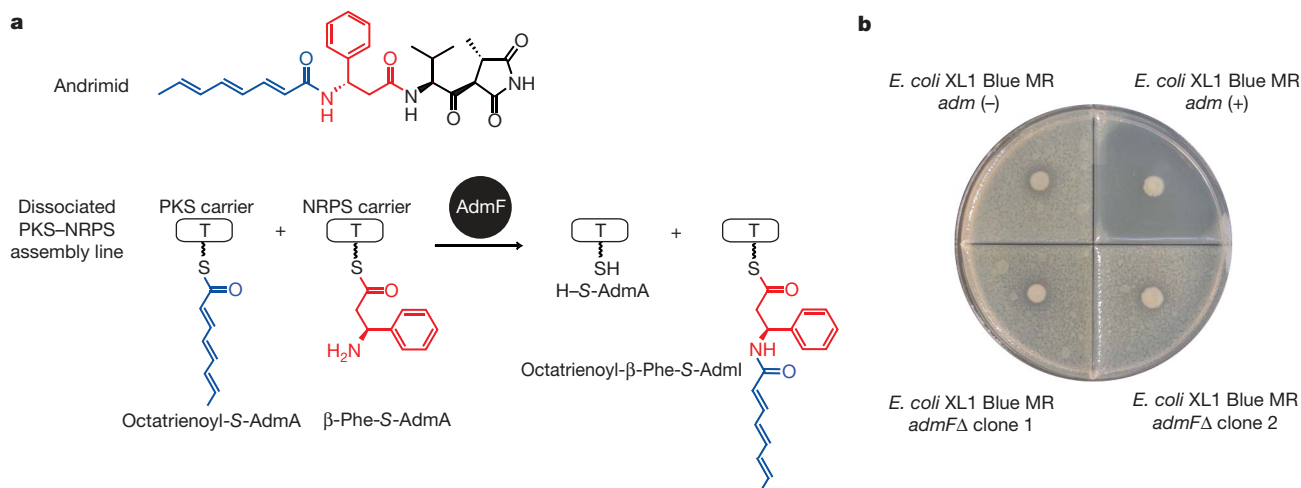


Figure 1 | First condensation step in andrimid biosynthesis. **a**, Structure of andrimid. The octatrienoyl chain and β -phenylalanine are shown in blue and red, respectively. AdmF, a transglutaminase homologue, catalyses the formation of an isopeptide bond between the octatrienoyl chain and the

amine group of β -phenylalanine. **b**, AdmF is required for andrimid production. Disruption of the *admF* gene in *E. coli* XL1 Blue MR (2194C1) abolishes the production of andrimid, as indicated by the absence of a zone of inhibition.

¹Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

the domains within these 'associated' mega-enzymes⁷. In contrast, the highly dissociated nature of the andrimid assembly line hinders prediction and demands biochemical experiments to delineate catalytic steps and assign functions to the separate domains responsible for activation and insertion of each building block in the andrimid backbone. For that reason, we first isolated the putative adenylation (A) domain containing NRPS enzymes AdmJ, AdmK and AdmP, and assayed their activity towards each amino acid in the pseudopeptide backbone of andrimid. More specifically, ATP-PP_i exchange assays determined that AdmK and AdmP, each made up of an A and T domain, activated L-valine and L-glycine to aminoacyl-AMPs, respectively (Supplementary Table 1). Another amino acid building block in andrimid, β -phenylalanine (β -Phe), was specifically activated by AdmJ, a stand-alone A domain. In fact, AdmJ demonstrated no detectable activity towards L-Phe, indicative of a high selectivity towards β -Phe over L-Phe (Supplementary Table 1). Immediately upstream of *admJ* in the *adm* cluster is *admI*, which encodes a free-standing T domain, which is most probably aminoacylated in its holo form by the action of its possible enzymatic partner, AdmJ. The AdmJ-dependent in *trans* aminoacylation of AdmI was first demonstrated by SDS-polyacrylamide gel electrophoresis (PAGE) and autoradiography using [¹⁴C] β -Phe as a substrate (results not shown). The AdmJ-catalysed aminoacylation of AdmI was confirmed using

non-radioactive β -Phe. The reaction mixture was analysed by quadrupole time-of-flight mass spectrometry (Q-TOF MS). The observed mass of 12,482.9 Da corresponded to β -Phe-S-AdmI (predicted mass 12,482.6 Da), thus authenticating the covalent attachment of β -Phe to the phosphopantetheinyl arm of holo-AdmI.

The biosynthesis of andrimid requires an intracellular pool of β -Phe. In the *adm* cluster, *admH* encodes for a protein homologous to the 4-methylidene-imidazol-5-one prosthetic group containing ammonia lyase EncP⁸ (63% identity) and aminomutase SgcC4 (36% identity; ref. 9). AdmH was expressed, purified and subsequently shown to catalyse the reversible formation of β -Phe from L-Phe with an equilibrium constant of 1.28 (Supplementary Fig. 2). No cinnamic acid was formed during the reaction, which suggested a highly coupled aminomutase reaction. In fact, reactions carried out using deuterated L-Phe (L-Phe-d₈) resulted in the exclusive formation of β -Phe-d₈, indicating the full migration of the deuterium atom from the C _{β} to C _{α} of the amino acid substrate. The efficient catalysis carried out by AdmH is in contrast with the activity of SgcC4 (ref. 9) and L-Phe aminomutase from *Taxus cuspidata* involved in taxol biosynthesis¹⁰, which both show significant uncoupling, as evidenced by the leakage of cinnamic acids during catalytic turnover.

These experiments show AdmI T domains are tethered with β -Phe in the early steps of andrimid formation. The role of the second

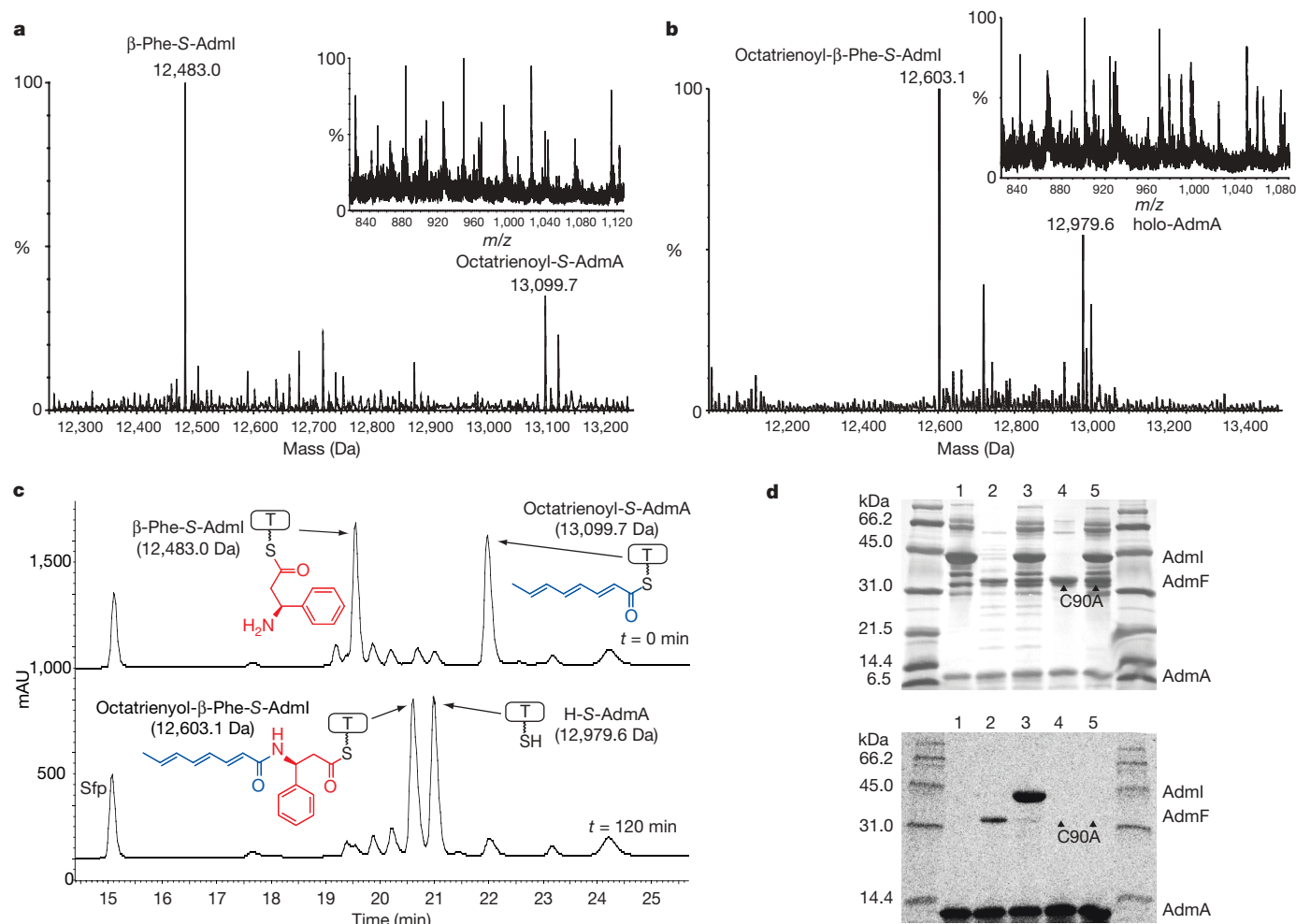


Figure 2 | Analysis of the AdmF-catalysed condensation of octatrienoyl to β -phenylalanine. **a**, **b**, Q-TOF mass spectra of the AdmF assays. **a**, Octatrienoyl-S-AdmA and β -Phe-S-AdmI. **b**, Octatrienoyl-S-AdmA and β -Phe-S-AdmI with AdmF. Refer to Supplementary Table 3 for interpretation. The TOF MS ES+ m/z spectrum used to calculate each mass spectrum is shown in the inset. **c**, HPLC trace of intact proteins obtained after incubation of octatrienoyl-S-AdmA, β -Phe-S-AdmI and AdmF. **d**, the

surrogate substrate [¹⁴C]butyryl-S-AdmA was prepared to visualize the AdmF-catalysed reaction. [¹⁴C]butyryl-S-AdmA was incubated with the GST-tagged β -Phe-S-AdmI (lane 1), AdmF (lane 2) or GST-tagged β -Phe-S-AdmI and AdmF (lane 3). [¹⁴C]butyryl-S-AdmA was incubated with AdmF C90A with or without GST-tagged β -Phe-S-AdmI (lanes 4 and 5). All reactions were quenched after 15 min incubations at 30 °C. Upper panel, SDS-PAGE; lower panel, autoradiogram. mAU, milli-absorbance units.

carrier protein, AdmA, associated with the first amide bond formed in andrimid, was more obvious because its gene clustered with the genes encoding a type II PKS for the octa-2,4,6-trienoic acid. AdmA, the only acyl-carrier-protein-like T domain encoded within the cluster, is therefore predicted to be the scaffold for creating this polyunsaturated fatty acid by the actions of the andrimid type II PKS enzymes (AdmC, AdmD and AdmE). The condensation between these two atypical biosynthetic units (that is, donor polyunsaturated fatty acid and acceptor β -amino acid) could not be readily explained, because no *in cis* or *in trans* C domain could be attributed to this portion of the andrimid pathway.

However, *admF*, present between *admA* and *admI*, drew attention because of possibly encoding a novel condensation candidate, as recently proposed². AdmF is predicted to possess a full Cys-His-Asp catalytic triad in a TGase core homologous to the one found in the structurally characterized Factor XIIIa TGase core: pfam01841 (ref. 11; Supplementary Fig. 3). A knockout of *admF* was prepared in a cosmid containing the otherwise complete *adm* cluster. *E. coli* XL1 Blue MR host cells heterologously expressing the complete cluster made the antibiotic, but *E. coli* expressing the *admFΔ* cluster were unable to produce andrimid (Fig. 1b). This observation established a requirement for AdmF and led us to investigate whether AdmF was indeed responsible for the *N*-acyl- β -peptide linkage in andrimid.

To interrogate the role of AdmF in andrimid biosynthesis, the early portion of the pathway was reconstituted *in vitro* (see Supplementary Fig. 4 for SDS-PAGE of corresponding proteins). Octatrienoyl-S-AdmA was prepared by direct loading of octa-2,4,6-trienoyl-CoA on apo-AdmA using the promiscuous phosphopantetheinyl transferase Sfp (see Methods).

The authenticity of the corresponding octatrienoyl-S-AdmA was confirmed by Q-TOF MS (Supplementary Table 2). The AdmF assays were carried out using octatrienoyl-S-AdmA and β -Phe-S-AdmI, prepared using AdmJ, as protein-tethered co-substrates. The reaction was monitored by analytical HPLC, in which the different proteins and their various states (that is, apo- (no phosphopantetheinyl arm), holo- (with phosphopantetheinyl arm), and substrate-loaded on phosphopantetheinyl arm) could be separated with baseline resolution (Supplementary Fig. 5). The AdmF reaction was characterized by the disappearance of two peaks with the concerted appearance of two new peaks, one of which co-eluted with a holo-AdmA standard. In control experiments where AdmF was omitted, no transfer reaction was observed (Supplementary Fig. 5). Further interrogation of reactions was accomplished by Q-TOF MS. The observed masses of the two products were 12,603.1 Da and 12,979.6 Da corresponding to octatrienoyl- β -Phe-S-AdmI (predicted mass 12,602.6) and holo-AdmA (predicted mass 12,979.8), respectively (Fig. 2a–c, and Supplementary Table 2 and Supplementary Fig. 6). The Q-TOF data confirmed that AdmF catalysed the transfer of the octatrienoyl chain from octatrienoyl-S-AdmA to β -Phe-S-AdmI, forming octatrienoyl- β -Phe-S-AdmI and holo-AdmA. AdmF did not require calcium for activity nor was it influenced by the presence of an excess of the chelating agent EDTA. The apparent turnover rate of AdmF was estimated to be $>40 \text{ min}^{-1}$ (Supplementary Fig. 8). Moreover, AdmF was not active towards untethered substrates, that is, free β -Phe and octatrienoyl-CoA. The AdmF reaction was repeated using β -Phe-d8-S-AdmI as acceptor substrate, which resulted in the predicted mass increase of 8 Da for the octatrienoyl- β -Phe-S-AdmI species and further confirmed that the octatrienoyl chain was being transferred to AdmI loaded with β -Phe (Supplementary Table 2 and Supplementary Fig. 7). To confirm amide bond formation, an alternative acyl donor was used to detect products resulting from the action of AdmF. AdmA was loaded with a ^{14}C -labelled butyryl group by using Sfp and ^{14}C -labelled butyryl-CoA. Following incubation with AdmF, T-domain-tethered intermediates were released by addition of base (see Methods). The identity of the AdmI-bound product as ^{14}C butyryl- β -Phe was shown by its co-elution with an authentic standard of butyryl- β -Phe (Supplementary Fig. 9).

Transglutaminases proceed through an acyl-enzyme intermediate involving an active site cysteine contained within a Cys-His-Asp catalytic triad³. Multiple sequence alignments suggested a corresponding catalytic triad in AdmF (Cys 90, His 126, Asp 143), predicting that Cys 90 is the nucleophilic residue responsible for the formation of an acyl-AdmF intermediate during catalysis (Supplementary Fig. 3). We used AdmA loaded with ^{14}C butyryl as a surrogate donor substrate to test this hypothesis. AdmF reactions performed using ^{14}C butyryl-S-AdmA resulted in the transfer of the ^{14}C butyryl chain to β -Phe-S-AdmI (expressed as a glutathione S-transferase (GST)-fusion protein for SDS-PAGE assays; Fig. 2d, lane 3). Importantly, AdmF itself became labelled during the reaction. A build up of the ^{14}C butyryl-S-AdmF intermediate was observed when the reaction was performed in the absence of the β -Phe-S-AdmI acceptor substrate (Fig. 2d, lane 2). The variant AdmF (C90A) was prepared to show that Cys 90 is the acylation site. The C90A mutation abolished both the activity of AdmF and its ability to form the acyl-enzyme intermediate (Fig. 2d, lanes 4 and 5), thus supporting the role of Cys 90 as a catalytic nucleophile in the AdmF-catalysed reaction.

Prior to this work, two structurally and mechanistically distinct condensation catalysts have been known to operate in natural product assembly lines: the KS domains of PKSs and the C domains of

Known antibiotic assembly lines

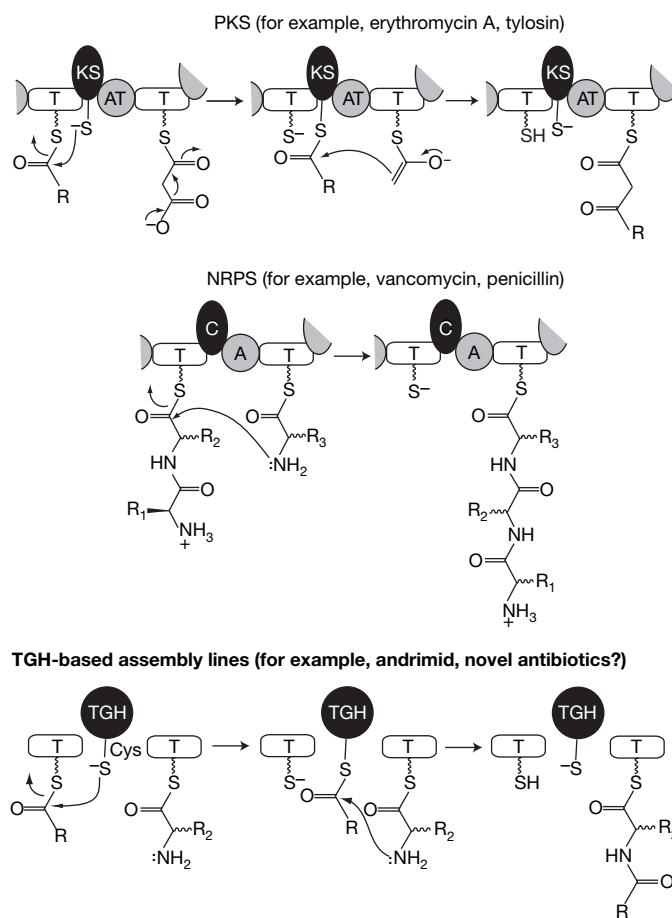


Figure 3 | Proposed mechanism of AdmF compared with prototypical ketosynthase (KS) domains in PKS and condensation (C) domains in NRPS. Top left panel, during PK biosynthesis, the KS domain catalyses a Claisen condensation to form a C–C bond and proceeds through an acyl-enzyme intermediate. Top right panel, during NRP biosynthesis, the C domain forms C–N bonds and does not proceed through an acyl-enzyme intermediate. Bottom panel, the transglutaminase homologue (TGH)-catalysed condensation proceeds through an acyl-enzyme intermediate and yields a C–N bond.

NRPSs. KS domains proceed through an acyl-enzyme intermediate to form carbon–carbon bonds between tethered acyl monomers (Fig. 3, top left), whereas C domains catalyse the formation of carbon–nitrogen bonds between tethered amino acid monomers¹². Unlike KS domains, C domains do not form acyl-enzyme intermediates during elongation (Fig. 3, top right)¹². We now reveal a third condensation catalyst used in antibiotic assembly, a TGase-homologue (TGH) that blends catalytic features of C and KS domains. AdmF serves as the prototype antibiotic assembly line TGH condensation catalyst and forms an acyl-enzyme intermediate through its active site cysteine en route to carbon–nitrogen bond formation (Fig. 3, bottom). Interestingly, a second TGH, AdmS, is present in the *adm* cluster. AdmS has extensive sequence identity with AdmF, and contains a putative active site Cys (Cys 95), indicating it would operate in a mechanistically related fashion (Supplementary Fig. 2). Insertionally inactivating *admS* abolishes andrimid production (Supplementary Fig. 10), proving it has a central activity in andrimid formation. AdmS is therefore the candidate to form a peptide bond between *N*-acyl- β -Phe and Val in andrimid.

Preliminary results indicate that AdmF is a promiscuous enzyme. In addition to octatrienoyl and butyryl acyl donors, AdmF also accepts crotonyl, hexanoyl and lauryl (results not shown), which heightens the value of this free-standing catalyst for combinatorial biosynthesis of new ‘unnatural’ antibiotics. Alternative β -aminoacyl-S-AdmI acceptors have yet to be tested. Searching the numerous sequenced bacterial genomes with AdmF and AdmS sequences identified a number of putative TGH domains encoded by genes found within orphan natural product biosynthetic clusters, notably within *Salinispora* (a marine actinobacteria) and *Photorhabdus* (a nematode symbiont)¹³. It is reasonable to suggest that TGH-containing natural product pathways could generate novel antibiotics active against drug-resistant bacterial pathogens.

METHODS SUMMARY

All proteins used in the current work were overexpressed in *E. coli*. The disruptions of *admF* and *admS* in cosmid 2194C1 were carried out using a gene-knockout procedure, as previously described¹⁴. ATP-PP_i exchange assays were performed using standard time course ATP/³²P-PP_i exchange assays followed by liquid scintillation counting. AdmA and AdmI were phosphopantetheinylated using Sfp and coenzyme A at 37 °C, following a procedure described previously¹⁵. Detailed procedures for AdmH and AdmF enzymatic assays are described in the Methods section of the current article.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 March; accepted 9 July 2007.

Published online 25 July 2007.

- Freiberg, C. *et al.* Identification and characterization of the first class of potent bacterial acetyl-CoA carboxylase inhibitors with antibacterial activity. *J. Biol. Chem.* **279**, 26066–26073 (2004).
- Jin, M., Fischbach, M. A. & Clardy, J. A biosynthetic gene cluster for the acetyl-CoA carboxylase inhibitor andrimid. *J. Am. Chem. Soc.* **128**, 10660–10661 (2006).

- Walsh, C. T. *Posttranslational modification of proteins, expanding nature's inventory* (Roberts and Company Publishers, Englewood, Colorado, 2006).
- Needham, J., Kelly, M. T., Ishige, M. & Andersen, R. J. Andrimid and moiramides a–c, metabolites produced in culture by a marine isolate of the bacterium *Pseudomonas fluorescens*: Structure elucidation and biosynthesis. *J. Org. Chem.* **59**, 2058–2063 (1994).
- Oclarit, J. M. *et al.* Anti-*Bacillus* substance in the marine sponge, *Hyattella* species, produced by an associated *Vibrio* species bacterium. *Microbios* **78**, 7–16 (1994).
- Fredenhagen, A. *et al.* Andrimid, a new peptide antibiotic produced by an intracellular bacterial symbiont isolated from a brown planthopper. *J. Am. Chem. Soc.* **109**, 4409–4411 (1987).
- Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).
- Xiang, L., Moore, B. & Inactivation, S. Complementation, and heterologous expression of *encP*, a novel bacterial phenylalanine ammonia-lyase gene. *J. Biol. Chem.* **277**, 32505–32509 (2002).
- Christenson, S. D., Wu, W., Spies, M. A., Shen, B. & Toney, M. D. Kinetic analysis of the 4-methylideneimidazole-5-one-containing tyrosine aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis. *Biochemistry* **42**, 12708–12718 (2003).
- Walker, K. D., Klettke, K., Akiyama, T. & Croteau, R. Cloning, heterologous expression, and characterization of a phenylalanine aminomutase involved in taxol biosynthesis. *J. Biol. Chem.* **279**, 53947–53954 (2004).
- Yee, V. C. *et al.* Three-dimensional structure of a transglutaminase: Human blood coagulation factor XIII. *Proc. Natl Acad. Sci. USA* **91**, 7296–7300 (1994).
- Keating, T. A. & Walsh, C. T. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Curr. Opin. Chem. Biol.* **3**, 598–606 (1999).
- Liolios, K., Tavernarakis, N., Hugenholtz, P. & Kyriakides, N. C. The genomes on line database (gold) v.2: A monitor of genome projects worldwide. *Nucleic Acids Res.* **34**, D332–D334 (2006).
- Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
- Quadri, L. E. N. *et al.* Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585–1595 (1998).
- Staab, H. A., Lueking, M. & Duerr, F. H. The preparation of imidazolides. Synthesis of amides, hydrazides, and hydroxamic acids by the imidazolidine method. *Chem. Ber.* **95**, 1275–1283 (1962).
- Tonge, P. J. *et al.* Localized electron polarization in a substrate analog binding to the active site of enoyl-CoA hydratase: Raman spectroscopic and conformational analyses of rotamers of hexadienoyl thioesters. *Biospectroscopy* **1**, 387–394 (1995).
- Sampson, B. A., Misra, R. & Benson, S. A. Identification and characterization of a new gene of *Escherichia coli* K-12 involved in outer membrane permeability. *Genetics* **122**, 491–501 (1989).
- Molnar-Perl, I. Derivatization and chromatographic behavior of the o-phthalaldehyde amino acid derivatives obtained with various SH-group-containing additives. *J. Chromatogr. A* **913**, 283–302 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Clardy for providing cosmid 2194C1 and M. Fischbach for helpful discussions. This work was supported in part by the NIH.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.T.W. (christopher_walsh@hms.harvard.edu).

METHODS

Chemicals. L-[¹⁴C]Phe, L-[¹⁴C]Val, L-[¹⁴C]Gly (each 250 mCi mmol⁻¹) and [¹⁴C]Butyryl-CoA were from American Radiolabelled Chemicals, [³²P]pyrophosphate was from Perkin Elmer Life Sciences, L-Phe-d8 was from Cambridge Isotopes, CoA and octa-2,4,6-trienoic acid and 1,1'-carbonyldiimidazole were from Sigma Aldrich. [¹⁴C]β-Phe and β-Phe-d8 were prepared enzymatically using AdmH (see 'AdmH activity assays' section). All other chemicals were of analytical grades.

Synthesis of octatrienoyl-CoA. The synthesis of octatrienoyl-CoA was adapted from a procedure developed for the synthesis of hexadienoyl-CoA^{16,17}. Equimolar amounts of octa-2,4,6-trienoic acid and carbonyldiimidazole were dissolved in dried THF at a final concentration of 0.2 M. The reaction mixture was incubated at room temperature for 6 h, as previously described. THF was evaporated and the resulting solid was re-dissolved in 1.1 ml of a solution of acetonitrile:0.1 M potassium phosphate buffer pH 7.0 (50:50). The mixture was incubated at room temperature for 15 min. Octatrienoyl-CoA was purified by HPLC on a VYDAC 218TP1022 (22 mm × 250 mm) column (GRACE VYDAC) using a water/methanol gradient (0 to 75% methanol over 45 min). The fractions containing octatrienoyl-CoA were pooled, lyophilized and stored at -20 °C. The authenticity of the octatrienoyl-CoA molecule was confirmed by ultraviolet-visible (UV-Vis) light spectroscopy and liquid chromatography-mass spectrometry.

Construction of plasmids. The genes *admA*, *admF*, *admH*, *admI*, *admJ*, *admK* and *AdmP* were amplified from the purified cosmid 2194C1 (ref. 2). The oligonucleotides used for PCR amplification are listed in Supplementary Table 3. The genes *admA*, *admF*, *admH* and *admI*, *admK*, *AdmP* were sub-cloned into pET28b, producing amino-terminal His-tagged fusions. *admI* and *admJ* were also sub-cloned into pET41a and pMAL-c2x, respectively, enabling expression of GST-AdmI and MalE-AdmJ fusion proteins. *E. coli* BL21 DE3 (Novagen) was used as the overexpression strain for all proteins. Site-directed mutagenesis of AdmF was performed using the Quick Change Mutagenesis kit (Stratagene), using oligonucleotides 'F_CtoA For' and 'F_CtoA Rev' (Supplementary Table 3). ***admF* knockout.** The *admF* knockout in the *adm* cluster was prepared in cosmid 2194C1 using the one-step inactivation by PCR products strategy¹⁴. The chloramphenicol resistance marker used to perform the *admF* knockout was amplified from plasmid pKD3 (ref. 14) using the following oligonucleotides:

admF-for-ko,

5'ATTCAACTCAATCAGATATTACGGCCTTAGGCGATGCAGCGTATACT-TATAGGAGGAAT-3';

admF-rev-ko,

5'CAAACCTCTGGCATGACTTGTACTCTCGTTGATAAGGATGGCACATCC-GACCTCGACGAAG-3'.

Andrimid bioassay. *E. coli* XL1 Blue MR (Stratagene) were transformed with cosmid 2194C1 or the corresponding *admF* deletion mutant derivatives. As a negative control, *E. coli* XL1 Blue MR was transformed with cosmid pSupercos1 (Stratagene). The resulting transformants were grown for 16 h at 37 °C in luria broth (LB) containing 100 μg ml⁻¹ ampicillin. Cultures (2.5 μl) were spotted on LB agar plates containing the same concentration of ampicillin. After 16 h of incubation at 30 °C, the plates were sprayed over with indicator strain *E. coli* imp⁻ (pKM079)¹⁸ and incubated for an additional 16 h at the same temperature. A clearing zone around the spotted cultures indicated production of andrimid.

Purification of proteins. His-tagged AdmA, AdmF, AdmH and AdmI were purified using Ni²⁺-nitrilotriacetate (Ni²⁺-NTA) agarose (Qiagen) followed by gel filtration on a Superdex 75 26/60 HiLoad (AdmA and AdmI) or Superdex 200 26/60 HiLoad (AdmH) or by anion-exchange chromatography on a monoQ 10/100 GL column (AdmF) (all columns from GE Healthcare). The GST-AdmI fusion protein was purified using a Glutathione Sepharose 4B resin (GE Healthcare). The MalE-AdmJ fusion was purified using an amylose resin (New England Biolabs) followed by gel filtration on a Superdex 200 26/60 HiLoad. The fractions corresponding to the non-tagged form of AdmJ (following cleavage at the Factor Xa protease site) were collected.

AdmH activity assays. AdmH aminomutase reactions were performed in 20 mM tricine, pH 8.25, at 30 °C in the presence of 1 mM L-Phe or β-Phe. The reaction was quenched at different time points by the addition of

o-phthalaldehyde-derivatization of the amino acids¹⁹. The quenched reactions were then analysed by HPLC on a Phenomenex Luna C18 reverse-phase chromatography column (4.6 × 250 mm) using an isocratic gradient of 40 mM sodium phosphate, pH 7.8:acetonitrile (75:25) 40 mM. The preparative reaction using L-[¹⁴C]Phe and L-Phe-d8 were performed without derivatization. The reaction mixtures were analysed and purified by HPLC using the column mentioned above and an isocratic gradient of methanol:water (50:50).

Acylation of AdmA. Acylation of AdmA was performed using the phosphopantetheinyl transferase Sfp and octatrienoyl-CoA of [¹⁴C]butyryl-CoA at 37 °C according to a published procedure¹⁵. Loading of AdmA with octatrienoyl-CoA was verified by Q-ToF MS. Loading of AdmA with [¹⁴C]butyryl-CoA was followed by autoradiography of the corresponding SDS-PAGE gel.

Adenylation domains amino acid selectivity assays. The activation of different amino acids by the A domains AdmJ, AdmK and AdmP were investigated using standard time course ATP/³²P_i exchange assays followed by liquid scintillation counting. AdmI was phosphopantetheinylated using Sfp and CoASH at 37 °C, following a procedure described previously¹⁵.

Aminoacylation assays for *in trans* loading of AdmI. The *in trans* aminoacylation of AdmI was investigated by incubating phosphopantetheinylated-AdmI with AdmJ, 1 mM ATP and 1 mM of L-[¹⁴C]Phe or β-[¹⁴C]Phe at 30 °C. The reactions were quenched at different time points by the addition of trichloroacetic acid. The protein pellets were resuspended and analysed by liquid scintillation counting. Aminoacylation of AdmI with unlabelled β-Phe was analysed by Q-ToF MS.

AdmF activity assays. The reaction catalysed by AdmF was investigated by incubating the enzyme with loaded AdmA and loaded AdmI in 50 mM Tris-Cl, pH 7.5, at 25 °C. For the HPLC assays, 50 μM octatrienoyl-S-AdmA and/or 50 μM β-Phe-S-AdmI were incubated with and without 1 μM AdmF. After a 60 min incubation at 25 °C, 50 μl of the reaction mixture was analysed on a Beckman System Gold HPLC system using a Phenomenex Jupiter C4 column (4.6 mm × 250 mm) and a 30 min 20–80% water–acetonitrile gradient. Identical reactions were carried out for Q-TOF MS analysis. Enzymatic assays for SDS-PAGE analysis were performed as follows: 50 μM [¹⁴C]butyryl-S-AdmA and 50 μM β-Phe-S-AdmI were incubated with and without 10 μM AdmF. The same reactions were performed using the AdmF Cys90Ala variant. SDS-PAGE gels were prepared by running 10 μl samples of the reaction mixtures. Autoradiograms of the resulting gels were collected using a Typhoon 9400 imager and analysed using the ImageQuant software (GE Healthcare). The activity of AdmF was also investigated using free CoA derivatives and free amino acids as substrates using concentrations of 250 μM and 1 mM, respectively. The nature of the bond formed between the acyl chain and β-Phe was investigated as follows. The reaction was carried out using 50 μM [¹⁴C]butyryl-S-AdmA, 50 μM β-Phe-S-AdmI and 1 μM AdmF in a total volume of 200 μl. The excess [¹⁴C]butyryl-CoA and β-Phe were removed by four consecutive washes with 400 μl of 20 mM HEPES, pH 7.5, in an Ultrafree filter microcentrifugal device (5 kDa cutoff; Millipore). The resulting protein solution was concentrated to 20 μl. Lithium hydroxide (80 μl of 0.1 N) was added and the mixture was incubated at 65 °C for 30 min. The released products were separated from the proteins by centrifuging the filter devices. The collected products were analysed by radio-HPLC and compared to authentic standards using a Beckman System Gold HPLC system equipped with a Beckman 171 radioisotope detector. The experiment was repeated with unlabelled substrates for LC-MS analyses.

Q-TOF mass spectrometry. Mass spectrometry experiments were performed on a Micromass Q-TOF 2 mass spectrometer coupled to a Waters 2690 liquid chromatography separation unit (Waters Corporation, Milford). Aliquots (10 μl) of protein samples were injected at a flow rate of 50 μl min⁻¹ using a water:acetonitrile (50:50) solvent mixture supplemented with 0.1% trifluoroacetic acid. The Micromass Q-TOF 2 instrument was calibrated using a phosphoric acid standard. The protein samples were analysed in the TOF MS positive electrospray (ES⁺) mode using a capillary voltage of 2,800 V and cone voltage of 30 V. The mass scan range was between 100 and 1,500 *m/z*. The resulting spectra were deconvoluted using the software MassEnt1, which yielded the calculated masses of the corresponding intact proteins.

LETTERS

Selection and evolution of enzymes from a partially randomized non-catalytic scaffold

Burckhard Seelig¹ & Jack W. Szostak¹

Enzymes are exceptional catalysts that facilitate a wide variety of reactions under mild conditions, achieving high rate-enhancements with excellent chemo-, regio- and stereoselectivities. There is considerable interest in developing new enzymes for the synthesis of chemicals and pharmaceuticals^{1–3} and as tools for molecular biology. Methods have been developed for modifying and improving existing enzymes through screening, selection and directed evolution^{4,5}. However, the design and evolution of truly novel enzymes has relied on extensive knowledge of the mechanism of the reaction^{6–10}. Here we show that genuinely new enzymatic activities can be created *de novo* without the need for prior mechanistic information by selection from a naive protein library of very high diversity, with product formation as the sole selection criterion. We used messenger RNA display, in which proteins are covalently linked to their encoding mRNA¹¹, to select for functional proteins from an *in vitro* translated protein library of $>10^{12}$ independent sequences without the constraints imposed by any *in vivo* step. This technique has been used to evolve new peptides and proteins that can bind a specific ligand^{12–18}, from both random-sequence libraries^{12,14–16} and libraries based on a known protein fold^{17,18}. We now describe the isolation of novel

RNA ligases from a library that is based on a zinc finger scaffold^{18,19}, followed by *in vitro* directed evolution to further optimize these enzymes. The resulting ligases exhibit multiple turnover with rate enhancements of more than two-million-fold.

We have devised a general scheme for the direct selection of enzymes catalysing bond-forming reactions from mRNA-displayed protein libraries (Fig. 1a). To demonstrate that new protein catalysts can be created using this scheme, we chose, as a model reaction, the ligation of two RNA molecules aligned on a template, with one RNA activated with a 5'-triphosphate (Fig. 1b). For our selection we used a library in which two loops of the small, stable human protein domain retinoid-X-receptor (RXR α) were randomized; this library has previously been used for the isolation of new ATP-binding proteins¹⁸ (Fig. 1c). We transcribed and translated this synthetic DNA library¹⁸ to generate mRNA-displayed proteins (Fig. 1a), which we then reverse-transcribed with a primer joined to the 3'-end of the 5'-triphosphorylated RNA (PPP-substrate, Fig. 1b). We incubated the library of 4×10^{12} unique mRNA-displayed proteins with the biotinylated oligonucleotide (HO-substrate) and the complementary splint oligonucleotide, which aligns the two substrate oligonucleotides. Proteins that catalysed the ligation of the two substrates

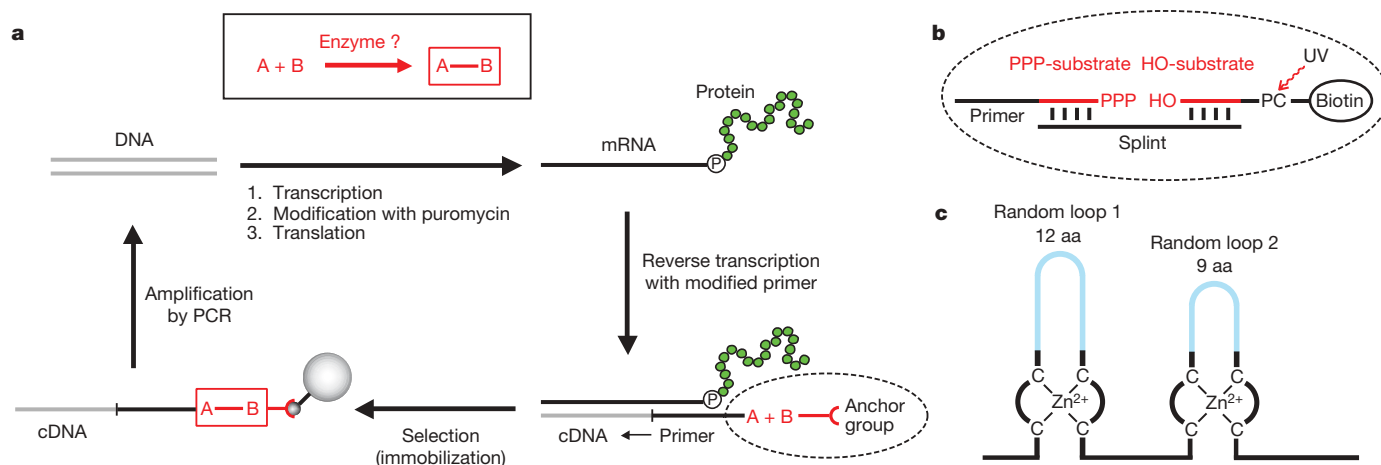


Figure 1 | *In vitro* selection of enzymes by mRNA display. **a**, General selection scheme for enzymes that catalyse bond-forming reactions. A DNA library is transcribed into RNA, cross-linked to a 3'-puromycin oligonucleotide, and *in vitro* translated. The library of mRNA-displayed proteins is reverse transcribed with a primer bearing substrate A. Substrate B, which carries an anchor group, is added. Proteins that join A and B attach the anchor group to their encoding cDNA. Selected cDNA sequences are then amplified by PCR, and used as input for the next round. **b**, Selection of enzymes that perform template-dependent ligation of

a 5'-triphosphate-activated RNA (PPP-substrate) to a second RNA with a 3'-hydroxyl (HO-substrate). The PPP-substrate is ligated to the primer and then used in the reverse transcription reaction. The cDNA of the catalytically active molecules is immobilized on streptavidin-coated beads via biotin, washed, and released by ultraviolet (UV)-irradiation of the photocleavable linker (PC). **c**, The scaffolded library¹⁸ is based on a two zinc finger domain with two loop regions (light blue) that are replaced by segments of 12 or 9 random amino acids (aa).

¹Howard Hughes Medical Institute, Department of Molecular Biology, and Center for Computational and Integrative Biology (CCIB), 7215 Simches Research Center, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA.

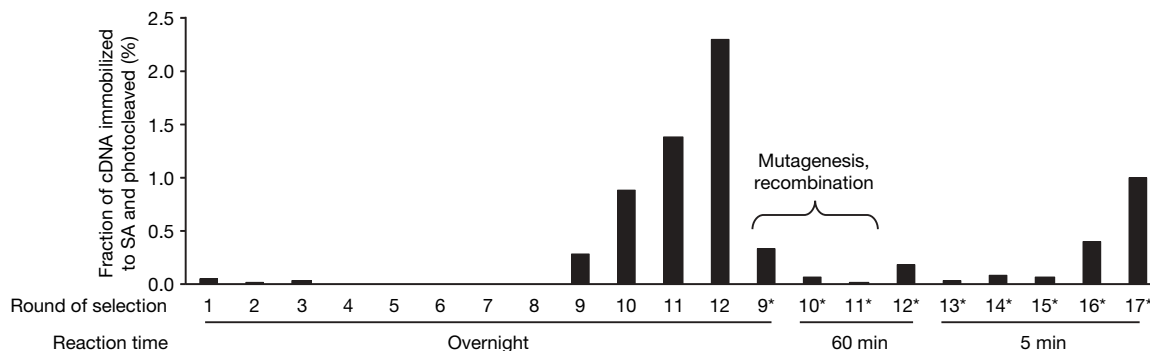


Figure 2 | Progress of the selection. The fraction of ^{32}P -labelled cDNA that bound to streptavidin agarose (SA) and eluted after photocleavage at each round of selection is shown. The input DNA into rounds 9*, 10* and 11* was subjected to mutagenic PCR amplification and, in addition, a recombination

procedure was performed before rounds 9* and 11*. The selection pressure was increased by decreasing the time of the reaction as indicated. Asterisks indicate selection rounds after mutagenesis and recombination.

covalently attached the biotin moiety to their own complementary DNA, which we captured on streptavidin-coated agarose beads. After washing, we eluted the cDNA by cleaving the photocleavable linker between the HO-substrate and the biotin. We amplified the cDNA by PCR and used it as input for the next round of selection and amplification. Over 9 rounds, the fraction of the input library immobilized on the streptavidin beads and then photoreleased increased from 0.01% to 0.3%, and after 12 rounds it increased to 2.3% (Fig. 2). To increase the activity of the selected ligases, we returned to the DNA library after round 8 and performed recombination and random mutagenesis^{20,21} by restriction enzyme digestion and ligation of the DNA and by subjecting the input DNA for rounds 9* through 11* (asterisks indicate selection rounds after mutagenesis and recombination) to error-prone PCR amplification. We then continued the cycles of selection and amplification without further recombination or mutagenesis until round 17, while increasing the selection pressure by gradually decreasing the reaction time from overnight to 5 min (Fig. 2).

The evolved pool of enzymes contained several families of closely related sequences as well as multiple unrelated single isolates (see Supplementary Information). Of the nine amino acids in loop 2, four positions were absolutely conserved in all sequences, four other sites were conserved in 86–90% of the clones and one position was conserved in 50% of the sequences. In contrast, we observed the motif DYKXXD at varying positions in the 12 originally randomized positions of loop 1 in 57% of the clones. This motif was probably enriched because it resembles the recognition site for the anti-Flag antibody M2 (ref. 22), which we used for purification of the mRNA-displayed proteins. These results indicate that the highly conserved loop 2 may have an essential role in ligase activity, whereas loop 1 may not be as important.

Analysis of the non-loop regions revealed a low conservation of specific cysteines of the original scaffold structure (Fig. 3). After 17 rounds of selection, just 16% of the clones (8 out of 49) retained the cysteine pattern as originally designed and were free of major deletions. The first and the fourth CX_nC sequences were highly conserved

(47 and 48 clones out of 49, respectively), but the second and third CX_nC motifs were retained in only 24% and 20% of the clones, respectively. In addition, two deletions (of 17 and 13 amino acids) were frequently observed (Fig. 3). Because of the mutation of up to half of the eight original zinc-coordinating cysteines and the frequent deletion of significant segments of the protein during the selection and evolution process, we believe that the majority of the proteins have undergone a substantial structural rearrangement in comparison to the original scaffold. We chose 18 clones from the final evolved library and screened them as mRNA-displayed proteins for ligation activity. All of the clones, including those with mutated cysteines or deleted regions, showed activity. We then expressed the 7 most active ligases (Fig. 3) in *Escherichia coli* as carboxy-terminal fusions with maltose binding protein (MBP) or without any fusion partner. All 7 enzymes were soluble when fused to MBP ($\geq 3 \text{ mg ml}^{-1}$ for several weeks at 4 °C). When expressed on their own, two of the ligases were soluble (6 and 7), whereas the other five precipitated or aggregated.

We chose the MBP-fusion of the most active enzyme (ligase 4) for more detailed characterization (Fig. 4). Incubation of the purified MBP-fusion enzyme with the PPP-substrate, the HO-substrate and the splint oligonucleotide yielded the desired ligation product (Fig. 4a, b) as well as the expected inorganic pyrophosphate by-product (Fig. 4c). We did not detect any product when we substituted the PPP-substrate with an oligonucleotide of identical sequence but with either a 5'-monophosphate or a 5'-hydroxyl instead of the 5'-triphosphate group (Fig. 4b). Preliminary experiments show that the enzyme catalyses the ligation equally well for all four nucleobases at the 3'-terminal base of the HO-substrate as long as they are correctly base-paired to the splint oligonucleotide. A mismatch at this position reduces the ligation efficiency several-fold. Enzymatic digestion of the ligated product confirmed the 3'–5' regiospecificity of the ligase reaction (Fig. 4d).

Because the RXR-library was based on a zinc finger protein, we examined the role of zinc and other cations in catalysis. The reaction

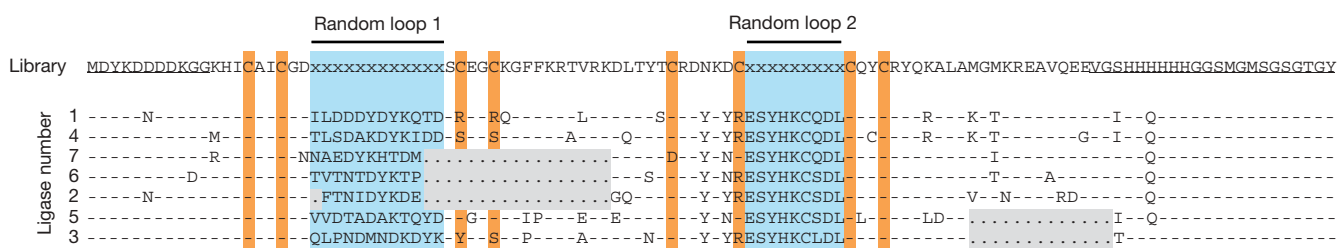


Figure 3 | Sequences of the starting library and selected ligases. Loop regions are highlighted in light blue. The cysteines highlighted in orange constitute the two pairs of CX_nC ($n = 2$ or 5) motifs that coordinate zinc ions in the original RXR α domain. Randomized amino acids in the library are

shown as x. Dashes indicate amino acids that are the same as in the starting library, whereas full-stops highlighted in grey symbolize deletions. The underlined flanking regions were not part of the RXR α domain but were added to form a Flag epitope tag, a hexahistidine tag and a linker region.

required Zn^{2+} and monovalent cations (K^+ or Na^+) with optima of 100 μM and 80 mM, respectively. The rate of the catalysed reaction showed a strong pH dependence with an optimum at pH 7.6. The optimal ligation conditions with regard to Zn^{2+} , monovalent cation and pH coincide with the conditions used during the selection. In contrast to the enzymatic reaction described here, the non-enzymatic template-directed ligation is inhibited by Zn^{2+} and shows a linear increase in reaction rate with increasing pH (ref. 23). Incubating the ligase with chelating resin (Chelex 100) resulted in an almost

complete loss of activity; activity could be restored by the addition of Zn^{2+} , but not by the addition of Cu^{2+} , Ni^{2+} , Co^{2+} , Mn^{2+} , Cd^{2+} or Mg^{2+} ions. Elemental analysis by inductively coupled plasma optical emission spectroscopy revealed 2.6 ± 0.4 equivalents (\pm s.d.) of bound zinc per ligase molecule, whereas the original wild-type RXR α protein contained 2.1 ± 0.1 molecules of zinc (\pm s.d.). The strong zinc dependence of the ligase enzymes could be due to either the continued existence of structural zinc sites or the presence of a catalytic zinc in the molecule.

To quantify the rate acceleration achieved by the selected ligase, we determined the rates of the catalysed as well as the uncatalysed RNA–RNA ligation reactions. We could not detect any uncatalysed formation of product in the absence of Mg^{2+} , consistent with previous work on the Mg^{2+} -dependence of the non-enzymatic ligation reaction²³. During the selection process, Mg^{2+} was present at a concentration of 5 mM, yet we found that the catalysed reaction did not require magnesium ions, and indeed was faster in its absence. By quantifying the detection limit of our assay we determined that the upper limit of the rate of the uncatalysed reaction of the pseudo-intramolecular complex of two substrate oligonucleotides pre-aligned on the template oligonucleotide in the absence of Mg^{2+} was $k_{\text{obs}} (\text{uncatalysed}) < 3 \times 10^{-7} \text{ h}^{-1}$. We measured the rate of the catalysed ligation in the absence of Mg^{2+} , at a subsaturating substrate concentration of 10 μM as $k_{\text{obs}} (\text{catalysed}) = 0.65 \pm 0.11 \text{ h}^{-1}$ (\pm s.d.), which is at least 2×10^6 -fold faster than the uncatalysed reaction. For the wild-type RXR α protein domain, we could not detect any ligated product (Fig. 4b).

We found that the evolved enzyme catalysed the ligation reaction with multiple turnover (Fig. 4e), although the selection scheme we employed used a single-turnover strategy that did not exert any selective pressure for product release. The intramolecular single-turnover design of the mRNA display selection scheme used here facilitates the isolation of enzymes, even if the rate acceleration is low or the substrate affinity is weak²⁴.

Preliminary biophysical studies suggest that the ligase possesses a folded structure. We chose ligase 6 for the following experiments because of its superior solubility in the absence of a fusion protein partner. Circular dichroism spectroscopy revealed an α -helical component of the secondary structure (Supplementary Fig. 1), and thermal denaturation indicated cooperative thermal unfolding (Fig. 4f). The two-dimensional $^1\text{H}/^{15}\text{N}$ heteronuclear single-quantum coherence (HSQC) NMR spectrum showed about 67 well-resolved peaks with a good chemical shift dispersion in the amide region of the spectrum, which indicates that a significant portion of the ligase protein is well folded (Supplementary Fig. 2a). A similar HSQC experiment with selectively ^{15}N -cysteine-labelled protein suggests that all six cysteines of ligase 6 are well structured (Supplementary Fig. 2b).

No natural enzyme is known to catalyse the ligation of a 5'-triphosphorylated RNA oligonucleotide to the terminal 3'-hydroxyl group of a second RNA, the reaction catalysed by the enzymes described here. An enzyme catalysing a similar reaction, the T4 RNA ligase, joins a 3'-hydroxyl group to a 5'-monophosphorylated RNA with the concomitant conversion of ATP to AMP and inorganic pyrophosphate via a covalent AMP-ligase intermediate²⁵. The reaction catalysed by the ligase described in this paper is more closely related to chain elongation by one nucleotide during RNA polymerization: in both cases, the growing strand and the triphosphate-containing substrate base pair to a template, the 3'-hydroxyl of the growing strand attacks the α -phosphate of a 5'-triphosphate, and a pyrophosphate is released in concert with the formation of a 3'-5' phosphodiester bond (Fig. 4a). RNA polymerases can be very fast, for example T7 RNA polymerase catalyses chain elongation at 240 nucleotides per second²⁶. Preliminary results with our selected ligase 4 did not show any polymerase activity with nucleoside triphosphates.

Ribozymes²⁷ and deoxyribozymes²⁸ previously selected from random oligonucleotide libraries catalyse the same reaction as our

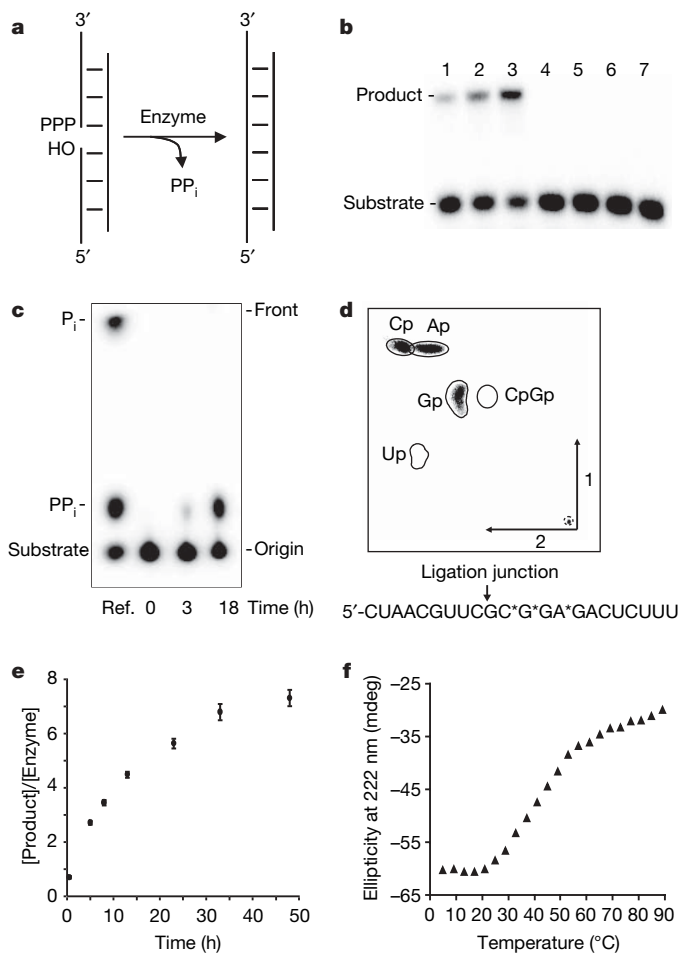


Figure 4 | Characterization of ligase enzyme. **a**, The RNA ligation reaction. **b**, Reaction catalysed by ligase 4 after 1, 3 and 10 h (lanes 1, 2, 3, respectively). Lanes 4–7, 10 h with no splint; 5'-monophosphate instead of PPP-substrate; 5'-hydroxyl instead of PPP-substrate; and wild-type RXR α protein domain instead of ligase 4. **c**, Release of inorganic pyrophosphate during ligation. Ligation reactions with γ - ^{32}P GTP-labelled PPP-substrate were separated by thin-layer chromatography. A mixture of inorganic ^{32}P -phosphate (P_i), ^{32}P -pyrophosphate (PP_i) and 5'- γ - ^{32}P -labelled PPP-substrate was run for reference (Ref.). **d**, 3'-5' regiospecificity of ligation. Ligation of α - ^{32}P GTP body-labelled PPP-substrate yielded product with ^{32}P at the indicated (*) positions. The product was digested to nucleoside monophosphates with RNase T2 (which does not efficiently digest 2'-5' RNA linkages) in the presence of a chemically synthesized RNA oligonucleotide that is identical in sequence to the predicted ligation product, but which contains a 2'-5' linkage at the ligation junction (5'-CUAACGUUCGC $^{2'}\text{p}^5$ GGAGACUCUUU). Digestion products were separated by two-dimensional thin-layer chromatography³⁰. Ultraviolet shadowing revealed the carrier RNA digestion products (Ap, Cp, Gp, Up), including the 2'-linked CpGp dinucleotide (encircled spots). Black spots represent the overlaid autoradiograph. The small dashed circle indicates the origin. **e**, Multiple turnover ligation. Substrate oligonucleotides and splint (each 20 μM) were incubated with ligase 4 (1 μM) for the indicated times and the ligation product was quantified. Error bars indicate s.d. **f**, Thermal unfolding of ligase 6 monitored by circular dichroism spectroscopy.

protein ligases. These ribozymes and deoxyribozymes have rate enhancements over the uncatalysed background reaction of the same order of magnitude as our protein ligase, and, in the case of the ribozymes, these rates were significantly increased by further *in vitro* evolution (up to 10^9 -fold rate acceleration)²⁹. Although our protein ligase was dependent on Zn^{2+} and inhibited by Mg^{2+} , the ribozyme-catalysed ligation is strongly dependent on Mg^{2+} with an optimum at $\sim 60 \text{ mM}$ ²⁷. The deoxyribozymes have been selected as Mg^{2+} -dependent variants and also as Zn^{2+} -dependent variants. The pH dependence of our ligase enzyme suggests that the catalytic mechanism involves acid–base catalysis by amino acid residues of the enzyme; in contrast, the pH-dependence of the ribozyme and deoxyribozyme ligases is more consistent with a catalytic role for one or more bound metal ions.

Our results represent the first use of mRNA display to select for a new enzyme activity. The general scheme could be readily adapted to a selection for other bond-making enzymes, and bond-breaking enzymes could be selected by immobilizing the mRNA-displayed proteins via the substrate to be cleaved and then enriching those molecules that liberate their encoding RNA. Alternatively, to isolate enzymes that catalyse other covalent modification reactions, the substrate could be attached to the cDNA and the active molecules separated with product-specific agents such as antibodies or aptamers. Thus, we believe that mRNA display represents a broadly applicable route to the isolation of novel enzymatic activities that are otherwise difficult to generate without explicit knowledge of structure or mechanism.

METHODS SUMMARY

Ligation activity assay of mRNA-displayed ligases by gel shift. Eighteen individual ligases were expressed separately as mRNA-displayed proteins, and incubated with HO-substrate and splint. After 5 h, the ligation reaction mixture was quenched, mixed with excess streptavidin, and separated by denaturing polyacrylamide gel electrophoresis (PAGE). The substrate and gel-shifted product bands were quantified.

Expression of ligases in *E. coli*. All proteins were expressed in Rosetta BL21 (DE3) cells and purified on either an amylose resin column (MBP-fused proteins) or a Ni-NTA resin column.

Ligation activity assay of free ligases. PPP-substrate ($20 \mu\text{M}$), $15 \mu\text{M}$ of splint and $10 \mu\text{M}$ of radiolabelled HO-substrate were incubated with $5 \mu\text{M}$ ligase, separated by PAGE and analysed. The k_{obs} values were determined by fitting the ratio of product concentration divided by enzyme concentration against time to a linear equation, and are the average of three independent experiments measured at less than 10% product formation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 February; accepted 19 June 2007.

- Schmid, A. *et al.* Industrial biocatalysis today and tomorrow. *Nature* **409**, 258–268 (2001).
- Cherry, J. R. & Fidantsef, A. L. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14**, 438–443 (2003).
- Panke, S., Held, M. & Wubbolts, M. Trends and innovations in industrial biocatalysis for the production of fine chemicals. *Curr. Opin. Biotechnol.* **15**, 272–279 (2004).
- Bloom, J. D. *et al.* Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **15**, 447–452 (2005).
- Aharoni, A., Griffiths, A. D. & Tawfik, D. S. High-throughput screens and selections of enzyme-encoding genes. *Curr. Opin. Chem. Biol.* **9**, 210–216 (2005).
- Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA* **98**, 14274–14279 (2001).
- Dwyer, M. A., Looger, L. L. & Hellinga, H. W. Computational design of a biologically active enzyme. *Science* **304**, 1967–1971 (2004).
- Park, H. S. *et al.* Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **311**, 535–538 (2006).

- Cesaro-Tadic, S. *et al.* Turnover-based *in vitro* selection and evolution of biocatalysts from a fully synthetic antibody library. *Nature Biotechnol.* **21**, 679–685 (2003).
- Xu, Y., Yamamoto, N. & Janda, K. D. Catalytic antibodies: hapten design strategies and screening methods. *Bioorg. Med. Chem.* **12**, 5247–5268 (2004).
- Roberts, R. W. & Szostak, J. W. RNA–peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl Acad. Sci. USA* **94**, 12297–12302 (1997).
- Keefe, A. D. & Szostak, J. W. Functional proteins from a random-sequence library. *Nature* **410**, 715–718 (2001).
- Wilson, D. S., Keefe, A. D. & Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl Acad. Sci. USA* **98**, 3750–3755 (2001).
- Baggio, R. *et al.* Identification of epitope-like consensus motifs using mRNA display. *J. Mol. Recognit.* **15**, 126–134 (2002).
- Cujec, T. P., Medeiros, P. F., Hammond, P., Rise, C. & Kreider, B. L. Selection of v-abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display. *Chem. Biol.* **9**, 253–264 (2002).
- Raffler, N. A., Schneider-Mergener, J. & Famulok, M. A novel class of small functional peptides that bind and inhibit human alpha-thrombin isolated by mRNA display. *Chem. Biol.* **10**, 69–79 (2003).
- Xu, L. H. *et al.* Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* **9**, 933–942 (2002).
- Cho, G. S. & Szostak, J. W. Directed evolution of ATP binding proteins from a zinc finger domain by using mRNA display. *Chem. Biol.* **13**, 139–147 (2006).
- Holmbeck, S. M. A. *et al.* High-resolution solution structure of the retinoid X receptor DNA-binding domain. *J. Mol. Biol.* **281**, 271–284 (1998).
- Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28–33 (1992).
- Wilson, D. S. & Keefe, A. D. In *Current Protocols in Molecular Biology* (eds Ausubel, F. *et al.*) Suppl. 51, unit 8.3 (Wiley, New York, 2000).
- Slootstra, J. W., Kuperus, D., Pluckthun, A. & Meloen, R. H. Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5. *Mol. Divers.* **2**, 156–164 (1997).
- Rohatgi, R., Bartel, D. P. & Szostak, J. W. Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J. Am. Chem. Soc.* **118**, 3332–3339 (1996).
- Griffiths, A. D. & Tawfik, D. S. Man-made enzymes — from design to *in vitro* compartmentalisation. *Curr. Opin. Biotechnol.* **11**, 338–353 (2000).
- Bryant, F. R. & Benkovic, S. J. On the mechanism of T4 RNA ligase. *J. Am. Chem. Soc.* **103**, 696–697 (1981).
- Bonner, G., Lafer, E. M. & Sousa, R. Characterization of a set of T7 RNA-polymerase active site mutants. *J. Biol. Chem.* **269**, 25120–25128 (1994).
- Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418 (1993).
- Purtha, W. E., Coppins, R. L., Smalley, M. K. & Silverman, S. K. General deoxyribozyme-catalyzed synthesis of native 3′–5′ RNA linkages. *J. Am. Chem. Soc.* **127**, 13124–13125 (2005).
- Eklund, E. H., Szostak, J. W. & Bartel, D. P. Structurally complex and highly-active RNA ligases derived from random RNA sequences. *Science* **269**, 364–370 (1995).
- Volckaert, G. & Fiers, W. Micro thin-layer techniques for rapid sequence-analysis of P-32-Labeled RNA - double digestion and pancreatic ribonuclease analyses. *Anal. Biochem.* **83**, 228–239 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. S. Cho for providing the RXR library before its publication along with valuable advice; A. D. Keefe for critical input during the initiation of this project; C. Mueller-Vahl for help with protein expression and purification; G.-P. Zhou and J. J. Chou for NMR measurements; A. D. Keefe, G. S. Cho, G. F. Short, R. Larralde, J. M. Carothers, J. K. Ichida, F. P. Seebeck, S. S. Mansy, C. Del Bianco, D. A. Treco, D. S. Wilson, A. J. Bell, A. Luptak, R. Bruckner and Z. Sachs for helpful discussions. This work was supported by a grant from the NASA Astrobiology Institute. B.S. was supported in part by the Emmy Noether-Programm of the Deutsche Forschungsgemeinschaft. J.W.S. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions All experiments were performed by B.S. Both authors designed the experiments, discussed the results and wrote the paper.

Author Information The DNA sequences encoding the ligase enzymes 1–7 have been deposited in GenBank under the accession numbers EU019543 to EU019549, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.W.S. (szostak@molbio.mgh.harvard.edu).

METHODS

Sequences of oligonucleotides are described in Supplementary Information.

Preparation of primer for reverse transcription. The reverse transcription primer was a chimaeric oligonucleotide made from a 5'-triphosphate RNA oligonucleotide and a DNA oligonucleotide at the 3'-end. The PPP-substrate (5'-pppGGAGACUCUUU) was synthesized by T7 RNA polymerase from a double-stranded template of BS47 and BS48 and purified by denaturing PAGE. The PPP-substrate was then ligated to BS50 in the presence of BS56 as template by T4 DNA ligase³¹ and the product was purified by denaturing PAGE to yield the reverse transcription primer: 5'-pppGGAGACUCUUUTTTTTTTTTTTTTTTCCAGATCCAGACATTC.

In vitro selection and evolution. The DNA library, designed and synthesized by Cho¹⁸, was PCR amplified with primers BS3long and BS24RXR2 to introduce a cross-link site at the 3'-end to use the psoralen-crosslinking protocol³² (for DNA library sequence see Supplementary Information). RNA was produced from the DNA library with T7 RNA polymerase. After purification by denaturing PAGE the RNA was photo-crosslinked³² with the XL-PSO oligonucleotide and ethanol precipitated. The mRNA-displayed proteins were generated as previously described^{11,12,18,33,34} with the following modifications. In the first round of selection a 10 ml translation was incubated at 30 °C for one hour (200 nM psoralen cross-linked RNA template, Red Nova Rabbit Reticulocyte Lysate (Novagen), used according to the manufacturer's instructions with an additional 100 mM KCl per 0.9 mM Mg(OAc)₂ and 69 nM ³⁵S-methionine). After addition of 450 mM KCl and 50 mM MgCl₂, the translation reaction was incubated at room temperature for 5 min and then diluted 10-fold into oligo(dT)cellulose binding buffer (10 mM EDTA, 1 M NaCl, 10 mM 2-mercaptoethanol, 20 mM Tris(hydroxymethyl) amino methane, pH 8.0, 0.2% w/v Triton X-100) and this mixture was incubated with 10 mg ml⁻¹ oligo(dT)cellulose (New England Biolabs) for 15 min at 4 °C with rotation. The oligo(dT)cellulose was washed on a chromatography column (Bio-Rad) with the same oligo(dT)cellulose binding buffer, then with oligo(dT)cellulose wash buffer (300 mM KCl, 5 mM 2-mercaptoethanol, 20 mM Tris(hydroxymethyl) amino methane, pH 8.0) and then eluted with oligo(dT)cellulose elution buffer (5 mM 2-mercaptoethanol, 2 mM Tris(hydroxymethyl) amino methane, pH 8.0) to yield 4 × 10¹³ mRNA-displayed proteins. The eluate was mixed with 10× Flag binding buffer (1× is 150 mM KCl, 5 mM 2-mercaptoethanol, 50 mM HEPES, pH 7.4, 0.01% w/v Triton X-100) and then incubated with 50 µl Anti-Flag M2-agarose affinity gel (Sigma; prewashed with Flag clean buffer (100 mM glycine, pH 3.5, 0.25% w/v Triton X-100) and Flag binding buffer) for 2 h at 4 °C with rotation. The Anti-Flag M2-agarose affinity gel was then washed with Flag binding buffer and eluted with Flag binding buffer containing two equivalents of Flag peptide (Sigma; one equivalent of Flag peptide saturates both antigen sites of the antibody resin) for 20 min at 4 °C with rotation. The eluate was diluted to a concentration of 5 nM mRNA-displayed proteins with an additional 50 mM Tris(hydroxymethyl) amino methane, pH 8.3, 3 mM MgCl₂, 10 mM 2-mercaptoethanol, 0.5 mM each of dCTP, dGTP, dTTP, 5 µM dATP, 50 nM α-³²P dATP, and used for the reverse transcription of the mRNA-displayed proteins with 50 nM reverse transcription primer and Superscript II (Gibco BRL) at 42 °C for 30 min. This sample was then dialysed twice against Flag binding buffer at a ratio of 1/1,000 and then incubated with 100 µl Anti-Flag M2-agarose affinity gel and processed as described for the first Flag affinity purification above. Zinc chloride and 5× selection buffer (1× is 400 mM KCl, 5 mM MgCl₂, 20 mM HEPES, pH 7.4, 0.01% w/v Triton X-100) was added to the Flag elution to make a final concentration of 100 µM and 1×, respectively. The mixture was incubated with 2 µM HO-substrate (PC-biotin) and 3 µM splint for the indicated times (Fig. 2) at room temperature. After quenching the reaction with 10 mM EDTA, the solution was incubated with 700 µl ImmunoPure immobilized streptavidin agarose (Pierce; prewashed with PBS buffer (138 mM NaCl, 2.7 mM KCl, 10 mM potassium phosphate, pH 7.4), including 2 mg ml⁻¹ tRNA (from baker's yeast, Sigma), and then washed with PBS alone) at room temperature for 20 min with rotation. The streptavidin (SA) agarose was washed on a chromatography column (Bio-Rad) with SA binding buffer (1 M NaCl, 10 mM HEPES, pH 7.2, 5 mM EDTA), with SA urea wash buffer (8 M urea, 0.1 M Tris(hydroxymethyl) amino methane, pH 7.4), with SA basic wash buffer (20 mM NaOH, 1 mM EDTA) and with water. For the first round of selection, the streptavidin agarose was used directly in the PCR amplification reaction (50 µl streptavidin agarose beads per 1 ml PCR). Every round was assayed by scintillation counting of the ³⁵S-methionine-labelled proteins (from translation to reverse transcription) or of the ³²P-labelled cDNA (after reverse transcription) to measure the efficiencies of the various steps. These data were then used to determine that the number of purified individual protein sequences introduced into the round 1 ligation reaction step (incubation with biotin-PC-RNA and splint RNA) was 4 × 10¹², based on the proportion of total methionine (translation) and total dATP (reverse transcription) incorporated

into the mRNA-displayed proteins, and the efficiency of each of the subsequent purification steps.

This procedure was repeated for 17 rounds except for the following changes: in round 2 and in all subsequent rounds the translation reaction was 2 ml, only 400 µl of streptavidin agarose were used and directly before the PCR amplification, the streptavidin agarose beads were aliquoted in a 50% PBS slurry to 100 µl open wells. The slurry was irradiated with a UV lamp (4 W) at 360 nm from a 1 cm distance for 15 min, while shaking to release the cDNA. The beads were filtered off and the solution was used for PCR amplification. Before round 9* and 11* the DNA was digested with restriction endonuclease *Ava*II, which recognizes a unique restriction site between the two zinc fingers, and then ligated back together with T4 DNA ligase to achieve a recombination of the two halves of the proteins. The input DNA in rounds 9*, 10* and 11* were further mutagenized by error prone PCR^{20,21} at an average mutagenic rate of 3.8% at the amino acid level.

Cloning. Cloning was done as in ref. 18 with some changes. To analyse the results of the selection, the cDNA of the respective round was cloned into the pCR-TOPO vector (TOPO TA Cloning) and the individual clones were sequenced. To express the proteins in *E. coli*, the ligase genes were amplified with primers BS63 and BS65, and the wild-type RXRα motif (courtesy of G. S. Cho) was amplified with primers BS68 and BS70. The PCR products were digested with *Nde*I and *Xho*I and cloned into the pIADL14 vector³⁵ (courtesy of I. Lessard and C. Walsh) to yield the MBP-fusion proteins or into the pET24a vector (Novagen) to yield the protein without any fusion partner.

Sequence analysis. For sequence alignments the following software was used: SeqLab of the GCG Wisconsin Package, BioEdit³⁶ and MultAlin³⁷.

Ligation activity assay of mRNA-displayed ligases by gel shift. The sequences of 18 individual ligases were amplified from their respective pCR-TOPO vector with primers BS3long/BS24RXR2 and separately subjected to one round of selection as described above. After the incubation with HO-substrate (PC-biotin) and splint for 5 h, the ligation reaction mixture was quenched with 10 mM EDTA/8 M urea and was then mixed with an excess of streptavidin (Pierce) and separated by denaturing PAGE. The gel was analysed using a GE Healthcare (Amersham Bioscience) phosphorimager and ImageQuant software.

Expression of ligases and wild-type RXRα in *E. coli*. All proteins were expressed in Rosetta BL21 (DE3) cells (Novagen) containing the recombinant plasmids at 37 °C in LB broth containing 50 µg ml⁻¹ kanamycin. Cells were harvested, resuspended in lysis buffer (400 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES, pH 7.5, 100 µM ZnCl₂, 10% glycerol) and sonicated. After centrifugation, the supernatant was applied to an amylose resin column (New England Biolabs) in the case of the MBP-fusion proteins. The immobilized protein was washed and then eluted with amylose elution buffer (150 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES, pH 7.5, 100 µM ZnCl₂, 10 mM maltose) and stored at 4 °C for further use.

To purify the proteins lacking the MBP-fusion the supernatant after centrifugation was applied to a Ni-NTA resin column (Qiagen) instead. The immobilized protein was washed and then eluted with acidic Ni-NTA elution buffer (20 mM NaOAc, pH 4.5, 400 mM NaCl, 5 mM 2-mercaptoethanol, 100 µM ZnCl₂) directly into a 1 M HEPES, pH 7.5 solution to yield a final concentration of 100 mM HEPES. For use in circular dichroism and NMR spectroscopy experiments, the protein was further purified by FPLC (BioCAD Sprint Perfusion System) using a Sephadex-200 gel filtration column (Pharmacia Biotech) with isocratic elution in 150 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES, pH 7.4, 100 µM ZnCl₂ at 4 °C. The proteins were stored at 4 °C for further use. Protein concentration was determined by the Bradford method.

Ligation activity assay of free ligases. PPP-substrate (20 µM; 11mer), 15 µM splint and 10 µM 5'-³²P-labelled HO-substrate (11mer) were incubated with 5 µM ligase in reaction buffer (100 mM NaCl, 20 mM HEPES, pH 7.5, 100 µM ZnCl₂) for the indicated time and separated and analysed as above.

Detection of pyrophosphate. The MBP fusion of ligase 4 (purified on an amylose column) was immobilized on Ni-NTA resin (Qiagen), washed with buffer (150 mM KCl, 5 mM 2-mercaptoethanol, 50 mM HEPES, pH 7.4, 0.01% w/v Triton X-100, 100 µM ZnCl₂) and eluted in acidic elution buffer (50 mM NaOAc, pH 4.5, 150 mM NaCl, 5 mM 2-mercaptoethanol, 100 µM ZnCl₂). The ligase was then dialysed against 150 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES, pH 7.5, 100 µM ZnCl₂. The ligase (3 µM) was incubated with 6 µM γ-³²P-labelled PPP-substrate (11mer), 9 µM splint and 12 µM HO-substrate (11mer). The reactions were separated by thin-layer chromatography on PEI cellulose plates, which were developed in 0.5 M KH₂PO₄ at pH 3.4.

Analysis of metal content. The MBP-fusion proteins of ligase 4 and wild-type RXRα (purified on amylose column) were dialysed three times against buffer (100 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES at pH 7.5; pre-treated with Chelex 100 beads (BioRad) for 3 h and filtered) at a ratio of 1/1,000. The metal content of 4 µM samples was measured with an Inductively Coupled Plasma Emission Spectrometer (Jarrell-Ash 965 ICP, University of Georgia).

Circular dichroism spectroscopy. Circular dichroism spectra were recorded on an Aviv circular dichroism Spectrometer Model 202. Wavelength scans were performed in 15 mM NaCl, 0.5 mM 2-mercaptoethanol, 2 mM HEPES at pH 7.5, 10 μ M ZnCl₂ and 100 μ M ligase 6 at 25 °C in a 0.1 mm cuvette at 1 nm bandwidth in 1 nm increments with an averaging time of 4 s. Thermal denaturation of 324 μ M ligase 6 in 150 mM NaCl, 5 mM 2-mercaptoethanol, 50 mM HEPES, pH 7.4, 100 μ M ZnCl₂ was monitored at 222 nm from 5 °C to 90 °C in 4 °C increments and an equilibration time of 2 min at each temperature step in a 1 mm cuvette at 1.5 nm bandwidth with an averaging time of 10 s.

NMR spectroscopy. ¹H¹⁵N-NMR spectra were recorded on Bruker 500 MHz and 600 MHz NMR instruments with either uniformly ¹⁵N-labelled or selectively ¹⁵N-cysteine labelled protein (0.3 mM) in 10% D₂O, 150 mM NaCl, 5 mM 2-mercaptoethanol, 50 mM HEPES, pH 7.4, 100 μ M ZnCl₂. Protein samples were prepared from minimal media cultures using ¹⁵N-labelled NH₄Cl as the sole source of nitrogen or ¹⁵N-labelled cysteine as the sole source of cysteine, respectively.

31. Moore, M. J. & Sharp, P. A. Site-specific modification of pre-messenger-RNA — the 2'-hydroxyl groups at the splice sites. *Science* **256**, 992–997 (1992).
32. Kurz, M., Gu, K. & Lohse, P. A. Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA–protein fusions. *Nucleic Acids Res.* **28**, e83 (2000).
33. Cho, G., Keefe, A. D., Liu, R. H., Wilson, D. S. & Szostak, J. W. Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. *J. Mol. Biol.* **297**, 309–319 (2000).
34. Liu, R. H., Barrick, J. E., Szostak, J. W. & Roberts, R. W. Optimized synthesis of RNA–protein fusions for *in vitro* protein selection. *Methods Enzymol.* **318**, 268–293 (2000).
35. McCafferty, D. G., Lessard, I. A. D. & Walsh, C. T. Mutational analysis of potential zinc-binding residues in the active site of the enterococcal D-Ala-D-Ala dipeptidase VanX. *Biochemistry* **36**, 10498–10505 (1997).
36. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**, 95–98 (1999).
37. Corpet, F. Multiple sequence alignment with hierarchical-clustering. *Nucleic Acids Res.* **16**, 10881–10890 (1988).

naturejobs

**JOBS OF
THE WEEK**

On 9 August, President George W. Bush signed into law a bill that could have broad implications for physicists, engineers and mathematicians across the United States. After two years of wrangling, the bill finally made it through the US Congress two weeks ago with strong bipartisan support for its provisions, which should boost funds and create programmes for science education and research.

The roots of the act lie in a 2005 report from the National Academies, which warned that the United States was falling behind other nations in maths and science. There's a degree of rhetoric in the title of the bill: the America COMPETES Act or, more fully, the America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education and Science Act, but there are several important provisions that should benefit the US scientific community.

The act recommends that the budgets for the National Science Foundation and the Department of Energy's Office of Science should double over seven years. And for the fiscal period 2008-10, it aims to give the former \$22 billion and the energy department as a whole \$17 billion. It also establishes a new Advanced Research Projects Agency for Energy, and provides \$150 million for education programmes in science, technology, engineering and mathematics. The act recommends that the budget for the National Institute of Standards and Technology should be doubled over ten years, and it establishes an early-career grant programme for researchers at universities and national labs doing work relevant to the energy department's mission, as well as a graduate-research fellowship programme.

The act seems to promise improved career and funding prospects for the physical and engineering sciences. But US lawmakers can be fickle. Scientists should not let policy changes affect their career choice until the money and programmes are in place. And, even then, they should proceed with caution — what's given can soon be taken away.

Gene Russo, acting editor of *Naturejobs*

CONTACTS

Acting Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
e-mail: a.douglas@nature.com
**Business Development
Manager:**
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:

Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Southwest UK/RoW:

Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

North Germany/The Netherlands:

Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development:
Tom Hancock

Naturejobs online production:

Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com



Positive outlook: the US national parks offer ideal conditions for a sabbatical star gazing.

The seven-year itch

Sabbaticals offer scientists respite and a chance at reinvention. **Hannah Hoag** explores the year off.

Tyler Nordgren is heading to Mount Desert Island in Maine — home of Acadia National Park. But it's neither the granite peaks nor the subalpine plant communities that have drawn him there. For Nordgren, a professor of physics at the University of Redlands in California, the attraction is the inky black night sky.

"The National Parks are some of the last dark-sky places in the United States where your average person can see the Milky Way," says Nordgren, who will spend a year as a nocturnal visitor to 12 national parks across the country. He'll deliver public lectures on astronomy, educate businesses about light pollution and work on a book connecting the parks to astronomy, planetary science and space exploration. "This is something I never could have done without a sabbatical," he says.

Like tenure, the sabbatical is a defining element of academic life that is almost unheard of in the corporate world. Although a sabbatical's features vary among institutions, most involve 6 to 12 months' leave from teaching and administrative duties every seven years. The break from the daily grind offers scientists time to reflect and recharge. But without careful planning the opportunity can easily be wasted.

For many scientists, the sabbatical comes soon after the stressful probationary years that lead to tenure. The prospect of free time can be exciting, but if it is not used constructively it becomes dead time, warns Nick Bampos, a chemist at the University of Cambridge, UK. "You're a bit like the kid in the candy store, eating too much and getting sick," he says. "You don't get any nutritional benefit."

In planning his six-month sabbatical, Bampos asked a trusted colleague to help him discuss his career, identify goals and prioritize the work he needed to finish. Rather than go abroad, he stayed at Cambridge,

reconnected with his research team, wrote papers and caught up on the work of others in his field. "I was able to look at my work in a more relaxed and analytical way that isn't available the rest of the time," he says.

Greg Scholes, a physical chemist at the University of Toronto, Canada, tapped a mentor's wisdom long before he took his sabbatical. Given the early stage of Scholes's career, the adviser suggested using the sabbatical to increase the depth of his knowledge and go deeper into problems, rather than, say, learning a new technique.

Scholes spent last summer at the National Renewable Energy Laboratory in Golden, Colorado — one of the many US national laboratories to host sabbatical scholars — learning about the way quantum dots and carbon nanotubes interact with light. He followed that up with nine months at the University of Pisa in Italy, working with collaborators on problems relating to the capture of light by the plant pigment chlorophyll. As a result, Scholes and his lab published more papers during his sabbatical than they had the previous year.

Time waits for no one

Bampos, Scholes and Nordgren all began planning their sabbaticals more than a year in advance.

"People tend not to plan far enough ahead," says Peter Petraitis, a marine ecologist at the University of Pennsylvania in Philadelphia, who completed his second sabbatical in January. Petraitis spent six months in Chile, where he lectured at the Pontifical Catholic University in Santiago, worked on a book, wrote a grant and travelled with his wife. "You need to have an honest, open conversation with yourself, your significant other and your kids. Planning in terms of your family is really important."

A sabbatical is a chance to do something outside the normal academic-year routine: write a book, take on a new project, work more closely with collaborators or experience research in an industrial or corporate setting. With some luck, the experience not only broadens horizons but advances a career.

During a typical academic year, Nordgren devotes most of his time to teaching undergraduates physics and astronomy. His research involves using large arrays of optical and radio telescopes to measure stars and the



Peter Petraitis says it helps to involve the family when planning a sabbatical.

diameter of the dark-matter haloes around merging galaxies. Nordgren considered bumping up his publication record and doing more research during his sabbatical. Instead, inspired by a recent visit to Yellowstone National Park, he took a risk. "The potential pay-off was wonderful," he says, "and I thought 'When else can I do this?'"

Next January, Albert Galy, a geochemist at the University of Cambridge, will be taking a long-overdue sabbatical at the University of Paris. He will work with his collaborator Jérôme Gaillardet on publishing the results from their 2000 expedition to Tibet. The pair will also move towards a more formal collaboration on the mechanisms of limestone weathering. But Galy will be gone for only six months. "I will be leaving graduate students and postdocs and starting two other PhD students in the autumn," he says.

Outside academia

Like most academics, Sam Roweis, a computer scientist at the University of Toronto, juggles the triple responsibilities of research, teaching and administration every day. He spends a good deal of his time designing computer algorithms that let machines look at huge amounts of data and pick out important patterns and relationships within those data. But for his sabbatical year, Roweis is in Google's 'search quality' department in Mountain View, California, where he is finding better ways to answer Internet users' questions.

"Going to a corporate research lab is a great way to make a big change," says Roweis. There he is able to focus on practical problems and address constraints such as speed and computer breakdowns, which don't exist in the academic world. "Google is the paragon of real-world constraints," he says. The corporate experience allows him to relate better to the graduate students who have had corporate jobs before graduate school.

But Roweis is also using his time at Google to make progress on a computer program that will help amateur astronomers read the night sky. The program — developed in collaboration with astronomers from New York University — can identify the stars captured in a photograph by searching through a star database. The project, which he hopes will become a sort of Google Earth for the stars, profits from the company's computing power and ability to search multiple databases quickly. "It's the type of project that really benefits from being at Google," he says.

Scientists who use their sabbaticals to get a fresh perspective can reinvent themselves. Anatomy professor Zena Werb has been at the University of California, San Francisco, for more than 30 years. She is now spending her third sabbatical as a visiting professor at the Max Planck Institute for Biochemistry in Martinsried, Germany, supported by an Alexander von Humboldt research award. "I wanted to think about what I want to do over the next ten years of my career, because they may be the last ten years when I am scientifically active," says Werb. The year-long break gives her some distance from committee meetings and requests to review papers.

Rather than spend the full year in Germany, Werb commutes from San Francisco in chunks of three to five weeks. The arrangement allows her to stay in close touch with her university, collaborate on new experiments with scientists at the Max Planck and work



Greg Scholes found that studying photosynthesis in sunny Italy boosted his lab's output.

Web links

Sabbatical in the Parks programme
www.nature.nps.gov/sabbaticals
 John Simon Guggenheim Memorial Foundation
www.gf.org
 Fulbright Awards
www.cies.org and www.fulbright.ca
 Alexander von Humboldt research awards
www.humboldt-foundation.de



Sam Roweis found that going to a corporate lab was a good way to make a big change.

on papers and grants uninterrupted. During the exchanges, which she began last September, she has become a mentor to the institute's graduate students, postdocs and junior faculty members. "We discuss their papers and projects," she says. "I'm not their boss; I won't make a judgement about their future."

Cost and benefits

Werb says she sees fewer and fewer people taking leave when it's due. "People aren't using enough of their time to be thoughtful," she says. "It's important to get the big picture instead of rushing around dealing with minutiae or the obvious." Werb blames flat funding, the inflating cost of doing science and the increased "care and feeding" of grants. "We spend an increasing amount of our time pursuing money," she says.

Fully paid sabbaticals are uncommon. Some universities will cover all or most of an academic's salary, but not the extra costs. Moving expenses, rentals and, sometimes, the loss of a second income mean that taking a sabbatical can mean taking a financial hit.

There are awards and fellowships that can top up a sabbatical income. Some scientists, including Petraitis, are supported by Fulbright Awards, which give academics from around the world the opportunity to do research or lecture in another country. Scientists from Latin America, the Caribbean, Canada and the United States can also apply for fellowships from the John Simon Guggenheim Memorial Foundation. The host institution, too, often provides a salary or stipend to visiting professors, as in Galy's and Werb's cases.

But worrying too much about funding can ruin the experience, cautions Werb. "Nothing," she says, "can replace the refreshment of a sabbatical."

Hannah Hoag is a freelance writer in Montreal.

Succussion

A radical solution.

Steve Longworth

We have just finished our retraining and now it's time to redeploy. Whoever would have thought it would come to this? We all started out with such high ideals. Well, except me of course. If I'm really honest, I only went into medicine because I thought it would be a great way for a rather ordinary looking bloke like me to meet lots of unattached, sexy young nurses who would then be sufficiently impressed by the title 'doctor' to form an orderly queue outside my bedroom door (and so it proved, I'm delighted to report). I guess that's why I've taken to this so easily. Ethics was never my strong suit. I'm really a rather cynical opportunist behind the carefully cultivated, charming, professional veneer.

Still, I'm apprehensive about our new role. I've never killed anyone before. Well, not intentionally (there was that rather unfortunate series of prescribing errors that the Trust swept under the carpet before hastily moving me on, but that's another story). Up to now, whenever I have stuck a knife into someone it was with their consent and under anaesthesia. This new role is going to take some getting used to, but I've always been open-minded. That's also why, despite the fierce scepticism and at times downright hostility of many of my colleagues, I was willing to use homeopathy. Let's face it, few other people were doing it and the general public are so gullible it meant that I could open up a nice little private practice and charge silly money for silly remedies with little competition.

You know the theory behind homeopathy? 'Like fights like'. So if you are treating a fever you give the patient something that causes a fever, but (and this is the important bit) you dilute it over and over again, and each time you dilute it the treatment gets stronger. With each dilution the solution is vigorously shaken, a process known as 'succussion'. There are those who speculate that succussion causes the water molecules to 'remember' the active ingredient even when Avogadro says there is not a trace of it left. Dilution makes it stronger. That's the part conventional science has the most trouble with, but there

are studies that show that homeopathic remedies do work, even in animals, so it can't all be the placebo effect despite the shaky theory. On the other hand I think that my bedside manner contributed just as much to the cures enjoyed by my pliant clientele as the elaborately prepared bottles of expensive water that I used to succuss. Success, suckers!

Perhaps we should have noticed sooner that something remarkable was happening. As life expectancy gradually rose throughout the early twentieth century we gave the credit to public-health reforms: clean water, efficient sewerage systems, the ending of unsanitary overcrowding and so on. When life expectancy continued to climb we pointed to our

is treated in a sewage farm and pumped offshore where it is diluted in the sea. The sea water evaporates to form clouds, with the water returning to earth as rain and collecting in reservoirs. We drink the water and the cycle starts all over again. Each step in the process involves a form of natural succussion. You see where this is going. With each dilution the original medicine gets stronger. Over 200 years every homeopathic remedy ever formulated has been diluted and blended over and over and over again, millions, maybe billions of times. At the end of the first decade of the twenty-first century a critical dilution threshold was reached. Now any glass of water from any tap anywhere in the world is the most powerful all-purpose drug ever created. No more cancer, infection, inflammation, degeneration; everyone will now live forever in perfect health, unless they are unlucky enough to suffer a sudden overwhelming physical injury. Immortality! What a disaster.

The world is rapidly overpopulating. There are simply not enough resources to go around. We may be immune to all known diseases and resistant to ageing but we still have to eat. House prices are stratospheric as every available space fills up. Water is the universal panacea and now universal panic's here. The world has been thoroughly shaken.

In India and the Far East, where population growth is the fastest, vast suicidal religious cults have emerged and are massively popular among the young and naive. Euthanasia for the Youth in Asia!

Here we have a solution based on our own massively popular mystical belief system — the National Lottery. Everyone has been allocated a unique set of numbers and the draw takes place twice weekly. It means we have been able to find important new work for all us unemployed doctors. We used to bring succour. Now no one wants a house call, but it's vital we force our way in, scalpel at the ready as we point the finger and proclaim the new Hypocritical Oath.

It's you!

Trust me, I'm a doctor.

Steve Longworth lives in Leicester, which was recently voted curry capital of Great Britain.



JACEY

increasingly powerful pharmacopoeia so that by the start of the third millennium just about everyone over 60 was taking a statin to lower their cholesterol and often a bagful of other prescribed drugs as well. But when we all became, to all intents and purposes, immortal, there had to be a radical new explanation.

Think about this. Homeopathy has been around for 200 years, and over that period every homeopathic remedy that has ever been formulated has been taken on countless occasions by millions of people. The potency increases with each dilution. So when you drink a homeopathic medicine it gets diluted in your total body water, then you pee it out and flush the toilet, so it gets diluted in the sewerage system. The sewage